

**SPATIAL STATISTICS FROM HYPERPLEXED
IMMUNOFLUORESCENCE IMAGES: TO
ELUCIDATE TUMOR MICROENVIRONMENT, TO
CHARACTERIZE INTRATUMOR
HETEROGENEITY, AND TO PREDICT
METASTATIC POTENTIAL**

by

Daniel M. Spagnolo

B.S. in Engineering, Swarthmore College, 2013

Submitted to the Graduate Faculty of
the School of Medicine in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Daniel M. Spagnolo

It was defended on

February 2, 2018

and approved by

D. Lansing Taylor PhD, Distinguished Professor of Computational & Systems Biology,

Director of Drug Discovery Institute

S. Chakra Chennubhotla PhD, Associate Professor of Computational & Systems Biology

Jeffrey Fine MD, Assistant Professor of Pathology, Magee-Womens Hospital of UPMC

Adrian V. Lee PhD, Professor of Pharmacology and Chemical Biology, Pittsburgh

Foundation Chair in Precision Medicine

Ge Yang PhD, Associate Professor of Biomedical Engineering and Computational Biology,

Carnegie Mellon University

Dissertation Advisors: D. Lansing Taylor PhD, Distinguished Professor of Computational

& Systems Biology, Director of Drug Discovery Institute,

S. Chakra Chennubhotla PhD, Associate Professor of Computational & Systems Biology

**SPATIAL STATISTICS FROM HYPERPLEXED
IMMUNOFLUORESCENCE IMAGES: TO ELUCIDATE TUMOR
MICROENVIRONMENT, TO CHARACTERIZE INTRATUMOR
HETEROGENEITY, AND TO PREDICT METASTATIC POTENTIAL**

Daniel M. Spagnolo, PhD

University of Pittsburgh, 2018

The composition of the tumor microenvironment (TME)—the malignant, immune, and stromal cells implicated in tumor biology as well as the extracellular matrix and noncellular elements—and the spatial relationships between its constituents are important diagnostic biomarkers for cancer progression, proliferation, and therapeutic response. In this thesis, we develop methods to quantify spatial intratumor heterogeneity (ITH). We apply a novel pattern recognition framework to phenotype cells, encode spatial information, and calculate pairwise association statistics between cell phenotypes in the tumor using pointwise mutual information. These association statistics are summarized in a heterogeneity map, used to compare and contrast cancer subtypes and identify interaction motifs that may underlie signaling pathways and functional heterogeneity.

Additionally, we test the prognostic power of spatial protein expression and association profiles for predicting clinical cancer staging and recurrence, using multivariate modeling techniques. By demonstrating the relationship between spatial ITH and outcome, we advocate this method as a novel source of information for cancer diagnostics. To this end, we have released an open-source analysis and visualization platform, THRIVE (Tumor Heterogeneity Research Image Visualization Environment), to segment and quantify multiplexed imaging samples, and assess underlying heterogeneity of those samples. The quantification of spatial ITH will uncover key spatial interactions, which contribute to disease proliferation and progression, and may confer metastatic potential in the primary neoplasm.

TABLE OF CONTENTS

PREFACE	x
1.0 INTRODUCTION	1
1.1 THE CHALLENGE OF SPATIAL HETEROGENEITY	1
1.2 HISTOPATHOLOMICS: AN OVERVIEW	4
1.3 MULTI-TO-HYPERPLEXED IMMUNOFLUORESCENCE	8
1.4 THESIS CONTRIBUTIONS AND OUTLINE	11
1.5 LIST OF PUBLICATIONS	14
2.0 VISUALIZATION OF SPATIAL INTRATUMOR HETEROGENEITY	15
2.1 ABSTRACT	15
2.2 INTRODUCTION	15
2.3 THRIVE PLATFORM DESCRIPTION	18
2.4 THRIVE CAPABILITIES	19
2.5 APPLICATION IN TUMOR HETEROGENEITY	21
2.6 CONTRIBUTIONS	22
2.7 ACKNOWLEDGEMENTS	23
3.0 QUANTIFICATION OF SPATIAL INTRATUMOR HETEROGENEITY IN BREAST CANCER	24
3.1 ABSTRACT	24
3.2 INTRODUCTION	25
3.3 METHODS	29
3.3.1 Tissue microarray preparation	29
3.3.2 Multiplexed immunofluorescence staining and imaging	29

3.3.3	Image processing and cell quantification	30
3.3.4	Data preparation and automated quality control	31
3.3.5	Partitioning cells into two subpopulations to account for the distinct regimes in biomarker intensity distributions	32
3.3.6	Learning dominant biomarker intensity pattern set from each cell sub- population	32
3.3.7	Determine best dictionary size, m	36
3.3.8	Construct spatial networks to describe the organization of biomarker patterns in a tumor	37
3.3.9	Using pointwise mutual information to quantify spatial biomarker pat- tern relationships	37
3.3.10	Visualize spatial networks for specific relationships	39
3.3.11	Software	39
3.4	DATASET	41
3.5	RESULTS	43
3.5.1	Preprocessing of multiplexed/hyperplexed immunofluorescence image data	43
3.5.2	A strategy for quantifying heterogeneity	43
3.5.3	Building a dictionary of biomarker expression patterns	45
3.5.4	Visualizing networks of spatial interactions from pointwise mutual in- formation maps	47
3.5.5	Pointwise mutual information maps as potential diagnostic biomarkers	48
3.6	DISCUSSION	51
3.6.1	Comparing pointwise mutual information to quadratic entropy and clinical standards	54
3.7	CONCLUSION	55
3.8	SUPPLEMENTARY MATERIALS	59
3.8.1	Information Theory & Machine Learning Glossary	59
3.8.2	Quadratic Entropy and Other Diversity Measures from Information Theory	61

3.9	ACKNOWLEDGEMENTS	62
4.0	SPATIAL ANALYSIS OF BIOMARKER EXPRESSION PATTERNS FOR PROGNOSTICATION OF OUTCOME IN COLON CANCER	64
4.1	ABSTRACT	64
4.2	INTRODUCTION	65
4.2.1	Colon cancer biology	65
4.2.2	Tumor microenvironment	66
4.3	DATA	68
4.3.1	Patient statistics	68
4.3.2	Imaging protocols and quality control	69
4.3.3	Experimental evidence of modeling potential	72
4.4	METHODS	76
4.4.1	Feature sets	79
4.4.2	Univariate modeling	80
4.4.3	Feature selection	81
4.4.4	Multivariate modeling	81
4.4.5	2-level cox model	82
4.5	RESULTS	82
4.5.1	Comparison of 11 models	82
4.5.2	Patient stratification	85
4.5.3	Spatial organization of risk, via quadrant analysis and cell neighbor- hood analysis	87
4.6	DISCUSSION	89
4.7	APPENDIX	93
4.7.1	Abbreviations	93
4.7.2	Colon cancer biomarkers	93
4.8	ACKNOWLEDGEMENTS	96
5.0	CONCLUSIONS	97
	BIBLIOGRAPHY	99

LIST OF TABLES

1	Tissue microarray (TMA) data with cohorts from various breast cancer subtypes.	42
2	Frequency table for cancer stage vs. cohort variables for complete patient cohort.	70
3	Frequency table for cancer stage vs. cohort variables for all chemotherapy-naïve patients.	71
4	Performance metrics on training data for prognostic feature sets.	86
5	Performance metrics on testing data for prognostic feature sets.	86
6	Colon cancer biomarkers.	94

LIST OF FIGURES

1	Depiction of the tumor microenvironment in the case of a carcinoma.	2
2	Quantifying spatial intratumor heterogeneity.	5
3	Data acquisition and image processing workflow for multiplexed immunofluorescence.	9
4	Tumor Heterogeneity Research Interactive Visualization Environment (THRIVE).	16
5	Multiplexed immunofluorescence image of a tissue microarray spot.	26
6	Canonical pointwise mutual information maps depicting various forms of spatial intratumor heterogeneity.	33
7	Learning a dictionary of cellular expression patterns.	34
8	Visualizing networks of spatial interactions from pointwise mutual information maps.	40
9	Pointwise mutual information maps as potential diagnostic biomarkers.	49
10	Quadratic entropy implementation with patient and core-level results.	53
11	Image segmentation results.	56
12	High and low signal subpopulation (L1 and L2) coefficient distributions.	57
13	ER(-) IDC distribution in comparison to other cancer cohorts.	58
14	Spatial homogeneity in colon cancer tissue microarray data under biomarker expression features.	73
15	Comparison of scatter plots for expression features vs. correlation features.	73
16	Comparison of t-distributed stochastic neighborhood embedding for expression features vs. correlation features.	74
17	Analysis workflow for recurrence prediction using biomarker features.	78

18	Kaplain-Meier Survival Curves for CRC 5-Year Recurrence Models.	88
19	Spatial organization of risk using correlation features.	90

PREFACE

I would like to acknowledge my two advisors, Lans and Chakra, for their mentorship and advice, not only in matters of graduate research but also in career development, goal setting, and personal growth. They went above and beyond in terms of making themselves available, believing in me when my confidence faltered, and supporting me on this journey. I'd also like to acknowledge the rest of my committee, Dr. Fine, Dr. Lee, and Dr. Yang, for helping me craft my aims by suggesting edits and sharing references. Without their help my aims may have taken me a decade to complete, rather than the half a decade that it actually took. Notably Dr. Rekha Gyanchandani in Dr. Lee's lab was of enormous help.

I am also lucky to have worked with many excellent scientists at General Electric Global Research. Weekly communications with these GE collaborators—Brion Sarachan, Yousef Al-Kofahi, Fiona Ginty, Dan Meyer, Chris Sevinsky, Peihong Zhu—have proven elucidating, and the projects we worked on together involved fascinating new forms of data.

My lab-mates, both past and present, have always provided key insights and friendly conversation. I'd like to acknowledge them here: Luong, Shikhar, Burak, Om, Maurice, Filippo, and Samantha. Additionally my graduate school cohort—Marta, Nick, Cihan, Kelvin, and Brad—thanks for your lunchtime conversations, and influencing me to rise to your level of excellence.

Also, I would be remiss if I didn't mention the support system of my friends and family. My parents always believed in my abilities, even when they couldn't quite describe what I was studying. My sisters were always a call or weekend visit away, and continue to inspire me with their intelligence and drive. My partner, MJ, spent many a day listening to me talk about science, and sneaking me healthy meals. My friend, James, probably listened to me complain more than anyone else. Thank you all.

1.0 INTRODUCTION

Tumor heterogeneity is observed in many different types of solid tumors and blood cancers [1]. In the thesis therein we define tumor heterogeneity as the presence of malignant cells within the tumor exhibiting varied morphology or phenotype, possibly as a consequence of clonal evolution. The topic of tumor heterogeneity has been burgeoning in recent years, under the hypothesis that heterogeneity may be used as a potential biomarker for cancer progression and response to therapy [2]. Furthermore, there are many efforts to better understand the effect of the cellular environment surrounding tumors, including the nearby immune cells, stromal cells, vasculature, and extracellular matrix. This so-called *tumor microenvironment* has already been shown to have an effect on tumor evolution [2, 3].

There are many challenges to overcome when describing the diversity inherent in heterogeneous tumors. Many different data types and analysis techniques are already in use, each with their own merits and demerits, with the potential for multi-platform integration. My work specifically uses bioimages of tissue microarray data to quantify heterogeneity and study its potential. In this chapter I will discuss why this problem is difficult and unique, and in particular why the spatial organization of malignant cells and the tumor microenvironment is important. Then I will discuss the way I will address these challenges, and where my research fits in this growing field. The following sections reflect these goals.

1.1 THE CHALLENGE OF SPATIAL HETEROGENEITY

It has long been observed that tumor cells exhibit phenotypic diversity. More recently evidence has suggested that tumors progress in a process akin to Darwinian evolution, con-

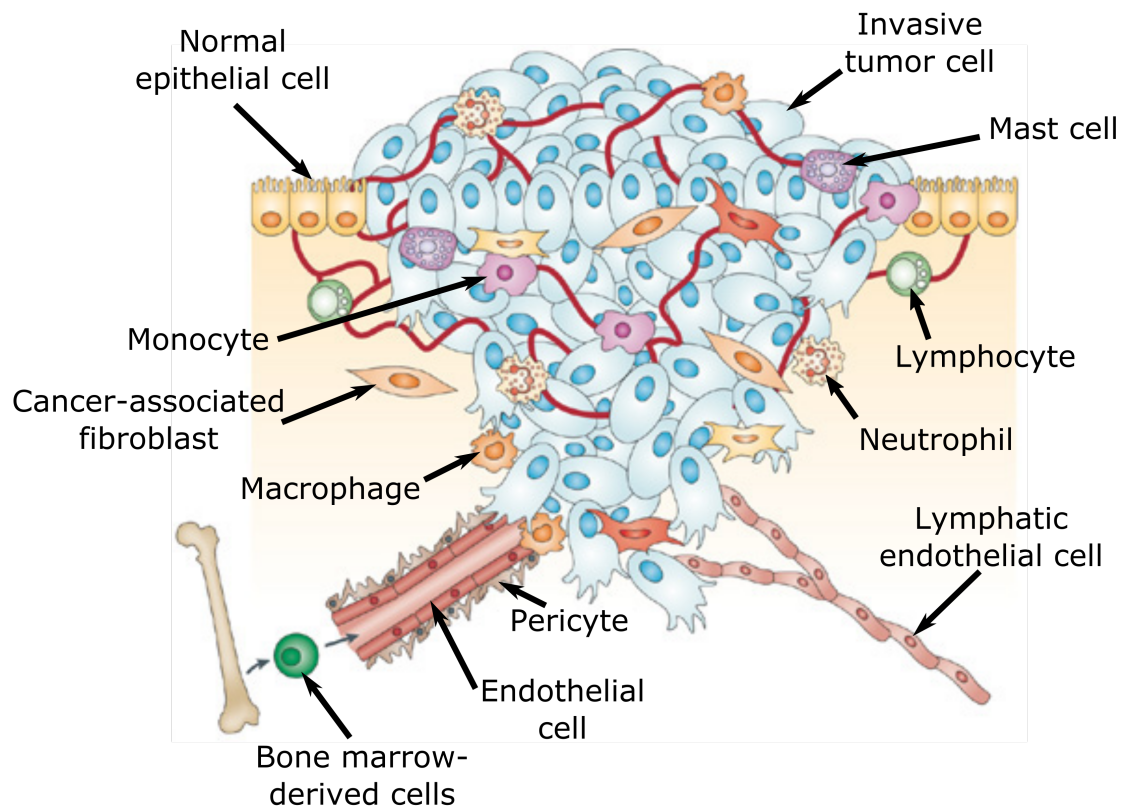


Figure 1: Depiction of the tumor microenvironment in the case of a carcinoma.

Tumor cells invade the surrounding epithelium and its environment, comprised of immune cells (e.g. lymphocytes, neutrophils, macrophages), stromal cells (e.g. fibroblasts), endothelial cells, and other co-conspirators. Adapted by permission from Springer Nature: Nature Reviews Cancer, Microenvironmental regulation of metastasis, Johanna A. Joyce, Jeffrey W. Pollard, 2009 [4].

tributing to the diversity of malignant cells within the tumor [5, 3]. Many envision tumor growth under a trunk-branch model, with the trunk representing ubiquitous driver events, and branches representing heterogeneous mutations, as opposed to more linear expansion models [6]. As a consequence of this progression, the resulting tumor may exhibit a high degree of *intratumor heterogeneity* (ITH), which can further promote tumor progression, enable drug resistance, and develop niches prone to metastasis. Intratumor heterogeneity, in this way, may impede the discovery of impactful biomarkers and the development of effective cancer therapies.

Intratumor heterogeneity as a phenomenon applies not only to the tumor cells, but also includes the tumor microenvironment (TME). Similar to how normal tissue is instructed to grow through paracrine signaling from neighboring cells or via systemic (endocrine) signals, human tumors likely also receive instructions in this cell-to-cell manner (although they can also generate their own growth signals) [2, 5]. In metastatic regions, as within the primary neoplasm, the TME can influence seeding of the cancer cells. A TME permissive to this seeding, through supportive stroma or other means, can become a “metastatic niche” in the body [7]. Figure 1 shows an example TME, including some of the potential cell types that may be implicated in this system. For example, vasculature is necessary for cell survival (via the nutrients it delivers), such that neoplasias must develop angiogenic ability to survive. Bone marrow-derived cells may help contribute to this ability. Cancer-associated fibroblasts may help build the extracellular matrix that supports advanced carcinomas. Immune cells can have either tumor-promoting or tumor-antagonising effects, depending on the circumstances. These are just a few examples of how the TME and tumor interact [5, 7]. Therefore the composition and spatial organization of the TME is as responsible as the subclonal composition of the tumor cells themselves for tumor progression and propensity of metastasis.

One of the many challenges encountered in the task of quantifying heterogeneity is a sampling problem. Often in the case of preparing a tumor biopsy in a formalin-fixed paraffin-embedded tissue block, cores are then extracted and assembled in a tissue microarray (TMA) for further analysis. Although there is “no standard approach to TMA creation, usage, or interpretation,” microarray cores are typically between 0.6-2.0mm in diameter [8]. Meanwhile,

whole tissue sections can be centimeters in length across. Thus, in an environment where heterogeneity exists not just in overall composition but is also spatially distributed across the tumor topology, it is possible that a given tumor core does not provide a complete story for studying the dynamics of the tumor as a whole. When using TMAs, we must always consider that our sampling may be imperfect. We must consider ways to compare and contrast tumor cores within a given patient. Finally, we must consider the way heterogeneous populations are spatially organized, as not all samples are created equally.

In the clinic, pathologists rely on phenotypic traits to stratify patients into biologically homogenous groups. The success of these classifications, and thus the potential for therapeutic benefit, is limited by the degree of intratumor heterogeneity. This is because the characteristics of the most abundant cell type may no longer be predictive of the tumor en masse [3]. Sampling can help in this regard, but not to the extent that I hypothesize accurate spatial models can assist in patient stratification and diagnosis. Intratumor heterogeneity also impedes drug delivery, and give tumors the ability to develop resistant phenotypes. Certain therapies may only target the dominant cell subpopulation [9], leaving behind quiescent tumor subpopulations which can then cause recurrence after the treatment period has ended. In the future, perhaps a measure for intratumor heterogeneity will be used by clinicians and the like to make informed patient decisions, as outlined in the hypothetical intelligent computer system of the 21st century digital pathology workflow [10].

In the following section we will discuss the current state of affairs for the study of intratumor heterogeneity.

1.2 HISTOPATHOLOMICS: AN OVERVIEW

Figure 2 outlines the array of information that can be used to characterize and quantify several factors that affect function or dynamics in normal-to-pathological human tissue samples. Here I define histopatholomics, a confluence of fields aspiring to understand the manifestations of disease, incorporating the microanatomy of tissue with the molecular elements of that tissue. These molecular elements include the genome (genes and noncoding DNA),

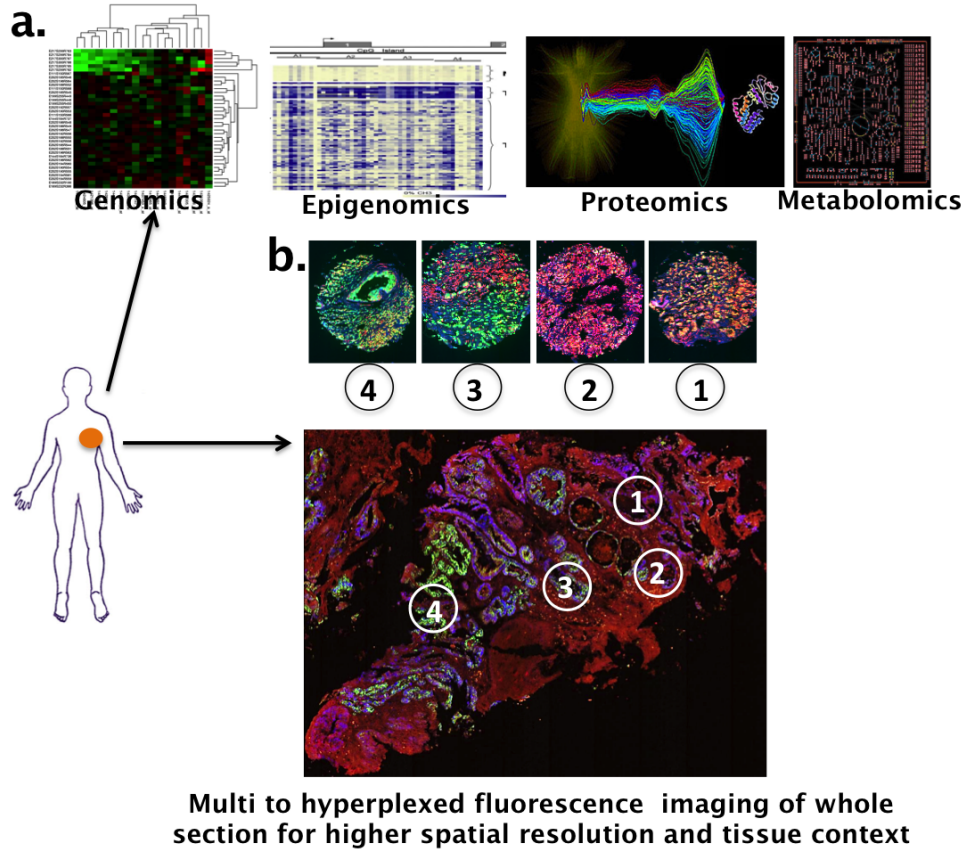


Figure 2: Quantifying spatial intratumor heterogeneity.

(a) The current state-of-the-art uses genomics, epigenomics, proteomics, and/or metabolomics to study heterogeneity on either ground up tissue samples or single cells. These approaches do not account for the spatial organization of the tumor microenvironment. (b) We present a method using multiplexed immunofluorescence imaging, which incorporates spatial distribution of biomarkers, in addition to their intensities, to characterize spatial intratumor heterogeneity. As shown in regions 1-4, the spatial organization of biomarker signals varies across the sample. Our method can capture this variation in a whole slide sample and is applicable to single-protein, multiplexed (up to 7 biomarkers), and hyperplexed (>7 biomarkers) immunofluorescence images. Furthermore, our method may be applied beyond the realm of cellular constituents, where we can study spatial interactions between cells and noncellular components (e.g., secretory elements, extracellular matrix)

epigenome (e.g. DNA methylation and histone modification patterns), metabolome, and the proteome. It is through this lens that I study spatial intratumor heterogeneity to elucidate spatial networks of proteins that may have unique consequences on cancer progression, metastasis, and/or response to therapy.

Several studies also seem to characterize cancers through this lens. Janiszewska et al. quantified intratumor heterogeneity of HER2 amplification and PIK3CA mutations in HER2-positive breast cancers, using STAR-FISH (specific-to-allele PCR-FISH), “a novel method for combined detection of single-nucleotide and copy number alterations in single cells in intact archived tissues” [11]. The sequencing of single cancer cells overcomes the challenge of detecting minor subclones (which may lead to therapy resistance) and predicting which mutations occur in individual cells, both of which hinder bulk tumor sequencing. With STAR-FISH, PCR products are visualized by hybridization with fluorescently labeled probes, allowing further study of the intratumoral topology. In patients with HER2-positive breast cancer undergoing neoadjuvant chemotherapy followed by adjuvant treatment with trastuzumab, a significant change in diversity across different areas of the same tumor (before and after chemotherapy) was associated with shorter disease free survival. The spatial dispersion of different cell types (wild-type PIK3CA signal, mutant PIK3CA signal, HER2 amplified cells, and all combinations) into clusters was also tracked pre- and post-treatment [11].

Another unique histopathologic approach to studying intratumor heterogeneity was presented by Gerlinger et al. Using exome sequencing, chromosome aberration analysis, and ploidy profiling on pretreatment biopsy of a primary renal neoplasm and chest-wall metastasis, as well as nine post-treatment primary-tumor regions and several post-treatment metastases, they revealed a branched evolutionary pattern across the treatment timeline. On average, a single biopsy revealed approximately 55% of all mutations detected in the tumor, indicating that single biopsies were not representative of the mutational landscape of the bulk tumor. They also determined that intratumor heterogeneity was not solely a consequence of the everolimus treatment. Additionally chromosomal aberrations were found to contribute to the genetic intratumor heterogeneity. Observing the mTOR pathway in their specimens, they were also able to link the functional heterogeneity of kinase activity to genetic intratumor heterogeneity [12].

Another multiregional sequencing approach by Hoefflin et al. combines precision quantitative imaging with regional whole-exome sequencing in primary clear cell renal cell carcinomas (ccRCC) [13]. One finding from this study was the discovery of intratumoral niches, notably in the tumor peripheral zone, which drive ITH via unique functional properties and mutations. Exome sequencing confirmed periphery-specific mutations, compared to the tumor center. They also found that for ccRCCs, functional ITH exists at the lowest staging, and does not increase with malignant progression. This implies that ITH is a feature of malignant growth, but not a necessarily a result of malignant progression.

Tirosch et al. used single-cell RNA sequencing (RNA-seq) to profile the cellular ecosystem of metastatic melanoma. By using t-distributed stochastic neighbor embedding and density clustering on the RNA-seq data, they “uncovered intra- and interindividual, spatial, functional, and genomic heterogeneity in melanoma cells and associated tumor components that shape the microenvironment, including immune cells, CAFs [cancer-associated fibroblasts], and endothelial cells” [14]. In doing so, they identified a dormant drug-resistant subpopulation, and discovered a subset of genes expressed by CAFs that influenced T cell proportion, which illustrates an example of intercellular communication effecting tumor phenotype. Similarly, Puram et al. use single-cell RNA-seq paired with immunohistochemistry to characterize the tumor microenvironment in head and neck cancer. Notably, they established partial epithelial-to-mesenchymal transition (p-EMT) as an independent predictor of nodal metastasis, grade, and other pathological features [15]. They found this p-EMT program to be spatially localized to the leading edge of primary tumors.

There exist many other studies in histopathomics, focusing on quantifying heterogeneity. While it may not be prudent to summarize them all here, I will briefly mention a few others of note. One popular approach involves measuring population averages from core samples of tumoral regions of interest via whole exome sequencing [16, 17, 18], epigenetics [19], proteomics [20, 21], or metabolomics [21]. Another approach uses single cell analyses, either by the methods listed above [22, 23], RNASeq [24], imaging [25], or flow cytometry [26]. A third approach uses light microscopy, which maintains the spatial organization of the TME, and uses molecular-specific labels for measuring the abundances of TME constituents [27, 28, 29].

While the methodology of each of these studies differs, the common thread lies in the holistic approach to studying intratumor heterogeneity, combining genetic and nongenetic factors with bioimaging, functional studies, and spatial information (including signal localization, cell-cell communication, and tissue architecture). The contributions of the tumor microenvironment have been found to be both tumor-antagonizing and tumor-promoting, and microenvironmental niches both in the primary neoplasm and metastases may be linked to tumor progression and treatment resistance. My research focuses on studies using multi-to-hyperplexed immunofluorescence, which has the benefit of single-cell resolution, spatial information, and potential for tracking a large number of tumor-affecting agents.

1.3 MULTI-TO-HYPERPLEXED IMMUNOFLUORESCENCE

Multiplexed immunofluorescence (MxIF) methods have the advantage of detecting several antigens (proteins or DNA probes, e.g.) simultaneously, while preserving the tumor architecture and the spatial organization of the individual cells and architectural elements [31]. This is essential for obtaining the morphological context of structures within the tissue (e.g. epithelium, vasculature, ducts and lobules, crypts and villi) [32]. With proper resolution and structural biomarkers, single-cell analysis can be achieved. MxIF methods, and other high-dimensional imaging methods, are uniquely positioned to model the tumor microenvironment and the spatial heterogeneity of this environment. Throughout this thesis, we will distinguish multiplexed from hyperplexed immunofluorescence by the image dimensionality (D), where each unique biomarker measured counts as an additional dimension. We delineate multiplexed images as 2D-to-7D images, and hyperplexed images as $> 7D$ images.

One MxIF method, developed by GE allows for the “subcellular characterization of multiple analytes in formalin-fixed, paraffin-embedded cancer tissue” [30]. Using chemical inactivation to quench fluorescent dyes after each image acquisition round, multi-to-hyperplexed images can be obtained using a co-staining of common dyes in iterative imaging cycles (see Figure 3). Images obtained from these iterative cycles are registered by aligning elements of a DAPI nuclear counterstain. Other structural biomarkers allow for cellular and subcellular

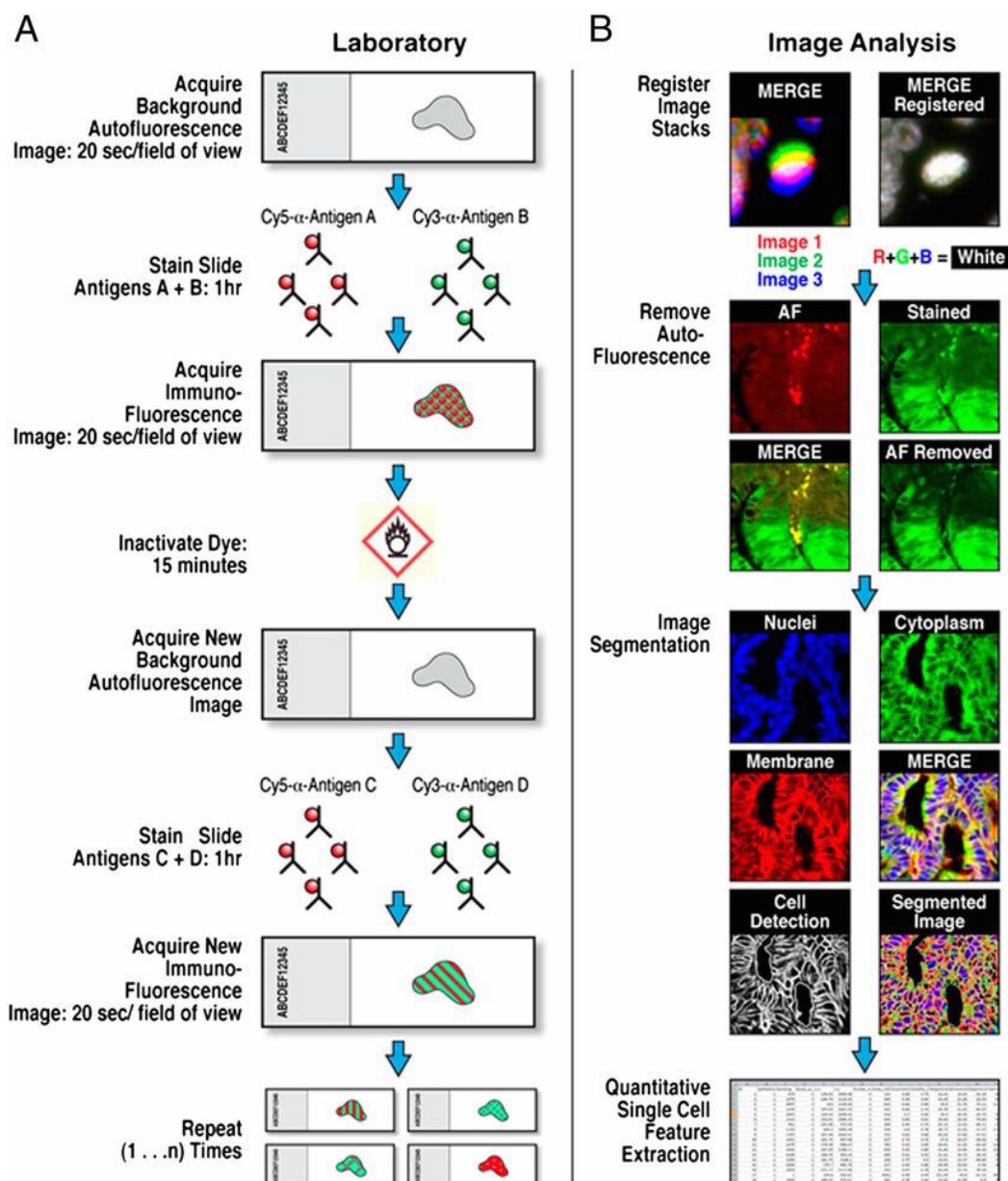


Figure 3: Data acquisition and image processing workflow for multiplexed immunofluorescence.

For the acquisition, the background immunofluorescence is captured, followed by images co-stained using Cy5 and Cy3. The co-stained dyes are quenched and restained for a different biomarker pair until the target antigen is exhausted. The stack of co-stained immunofluorescence images is registered, with autofluorescence removed, before single cell segmentation and quantification is applied. Adapted by permission from the Proceedings of the National Academy of Sciences of the United States of America: Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue, Michael J. Gerdes et al., 2013 [30].

resolution within the tissue samples. Currently, this method allows for the quantification 60+ analytes, with minimal loss of target epitopes or tissue integrity. The novelty of this method lies in the development and validation of a fluorophore inactivation solution, the ability for combined analysis of nucleic acids and proteins from the same tissue sample, integration with histological stains, and subcellular quantitation of highly multiplexed data [30].

The GE MxIF method has been adopted by many research labs in recent years. Nelson et al. used this technology to study organ morphogenesis in the submandibular salivary gland (SMG). By using single cell resolution to track 20 epithelial progenitor and differentiation markers, they were able to quantify the spatiotemporal progression of multiple progenitor cell populations and their role in SMG development and tissue homeostasis [33]. Clarke et al. validated the GE MxIF method, showing concordance between IHC scoring and the fluorescence data using the correlation scores of standard breast cancer biomarkers [32]. Uhlik et al., in a radiogenomic study, discover distinctive stromal phenotypes for new therapeutic hypothesis using RNA expression profiles, multiplexed IHC, and GE MxIF for vascular and immune cell status [34]. Gerdes et al. also use this technology to study ductal heterogeneity in carcinoma in situ of breast, using both epithelial and immune/stromal components [35].

While the GE MxIF method is perhaps the latest multiplexed fluorescence technology to gain traction, other platforms have been developed to similarly map larger quantities of analytes in tissues sections than standard multi-channel immunofluorescence. The Toponome Imaging System (TIS) precedes GE MxIF by several years, leveraging Schubert et al.’s multiple-epitope-ligand cartography (MELC) technology to “map the location of several proteins in the one sample [sic] of cells or tissue using sequential rounds of fluorescent detection in situ” [36, 37]. It claims the capability of imaging over 100 different molecular components at continuous intensity levels, to GE’s currently validated 60+ analytes (although GE claims the upper limit of analytes in a single MxIF assay has not yet been reached) [30, 36]. The main discernable difference in the methodology between the two methods is in the fluorescence deactivation step, where MELC uses photobleaching via soft multiwavelength excitation and GE uses chemical inactivation. Additionally, MELC is not compatible with formalin-fixed paraffin-embedded tissue [32]. There is some debate as to whether photobleaching (TIS) or chemical inactivation (GE) results in a larger signal loss

of fluorophores (and thus poorer signal-to-noise ratio) across staining rounds, and similarly there are discussions about the reproducibility of both systems [32, 38, 39]. TIS has been used to find an ALS motif—a spatial network code of proteins—surrounding a lead-protein CD16, which can be used for early therapeutic decisions, and was confirmed clinically [40].

Another method, multiplexed ion-beam imaging (MIBI), also touts an 100 target limit for simultaneous imaging [41]. This method uses secondary ion mass spectrometry to image antibodies to circumvent spectral overlapping and other limitations encountered in fluorescence-based flow cytometry, however remains burdened by throughput issues. Rost et al. show a strong correlation between HER2 quantitation by MIBI analysis and pathologist derived HER2 scoring in breast carcinoma tissue [42]. Lin et al.’s cyclic immunofluorescence (Cy-cIF) claims to present a public-domain method for achieving high multiplicity using standard reagents [43]. One difference I have noted is that CycIF seems to bleach all dyes in sequential rounds, while GE MxIF leaves a control dye untouched across quenching rounds.

Overall, these imaging methods show much promise but still have many barriers to entry. The cost of instruments and reagents can be prohibitive, image scanning can be too slow for large samples, reproducibility can be challenging from lab-to-lab, and “complete removal of antibodies often requires treatments deleterious to the sample” [31]. Additionally, there are no standardized protocols for preprocessing multiplexed images, either via image normalization across imaging slides, accounting for fluorophore signal loss, or controlling the varying dynamic range of fluorescence across multiple antigens. Yet, much like the early ages of DNA sequencing, there are bound to be obstacles and a certain degree of uncertainty on the route to revolutionary forms of data for biological research.

1.4 THESIS CONTRIBUTIONS AND OUTLINE

The use of multiplexed immunofluorescence, and other high-dimensional imaging modalities, facilitates unprecedented insights into disease states via protein-protein interactions and localizations [44, 45]. In this thesis, I study the potential for discovering molecular ITH through these imaging modalities. An area I seek to innovate is the encoding of spatial

information of the TME into ITH metrics. I hypothesize that measures of spatial ITH will provide unique insights in the study of disease progression and response to treatment. Incorporating the TME is of utmost importance because signaling between the cells in the TME allows tissue to achieve phenotypes that cannot be achieved by any cell type in isolation. In this way, we view cancer as an emergent heterocellular phenotype [46].

The first contribution of this thesis is to present a novel method to quantify spatial intratumor heterogeneity. Our method elucidates the spatial relationships between TME constituents, which can be preferentially expressed in different tumor architectural elements or biological processes. In our approach, each cell is phenotyped by its biomarker intensity vector, using a sparse overcomplete pattern recognition paradigm. Pointwise mutual information (PMI), computed over the spatial cellular network, uncovers key spatial interactions between cell phenotypes in the tumor sample. PMI scores are compared and contrasted for patients belonging to several disease cohorts, as defined by pathologists. PMI, as a spatial heterogeneity measure, is compared to previous heterogeneity quantification approaches, such as quadratic entropy (QE), Shannon entropy, and Simpson index. Spatial ITH poses a major challenge for cancer treatment since tumor biopsies used to inform diagnosis may not be representative of the tumor at large, and thus may skew decision-making processes [47, 48]. Furthermore, the spatial complexity of the TME—the interactions between the TME constituents—may contribute significantly to treatment resistance [49]. By characterizing spatial ITH, we are providing novel approaches for understanding spatial relationships in the TME, improving accessibility of multiplexed imaging data, and potentially providing a key element of future cancer diagnostic tools.

The second contribution of this thesis is to test measures of the spatial cell-cell communications between TME constituents with correlation to clinical outcome (e.g. cancer recurrence, staging, progression to metastasis). Since spatial ITH impedes therapy, the degree of ITH and the spatial composition of the TME is found to affect the progression of disease. A variety of spatial metrics are tested for prognostic potential using univariate survival models, and a robust combination of spatial features is shown to affect outcome using a multivariate Cox Proportional Hazards model. Efforts are also made to interrogate how disease risk is spatially organized within tissue microarray cores. Kaplan-Meier curves

show how spatial metrics can be used to segregate a patient cohort into low and high risk groups, which can assist in the development of personalized cancer patient management. If we find a strong association between spatial heterogeneity and outcome, then we can provide clinicians with previously inaccessible information related to disease progression and proliferation. Additionally we can use these statistical tests to identify biomarker panels ideally suited for their prognostic ability in the clinic.

The culmination of these contributions is the development and deployment of THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), a software suite for interacting with multiplexed/hyperplexed IF images and visualizing spatial ITH. THRIVE displays ITH data in a meaningful format, allowing end users to visualize key features of ITH that may evolve as prognostic biomarkers of progression and therapeutic response and point to underlying molecular mechanisms. Soon, THRIVE will include functionality to identify potential associations between heterogeneity and clinical features of interest. An open-source prototype has already been disseminated to the cancer research community, and a mechanism by which others can contribute capabilities, including new algorithms and visualizations, will be added to the next update. The main goal of THRIVE is to reduce intra- and inter-observer variability in quantifying heterogeneity at diagnostic laboratories, and to enable a variety of heterogeneity quantification methods to be evaluated and compared in a consistent framework. THRIVE is a user-friendly interface to a robust and flexible analysis engine that will co-evolve with imaging technology. This is a desperate need to advance cancer research.

Overall, my work leverages high-dimensional imaging modalities to quantify spatial intratumor heterogeneity, and demonstrate the prognostic power of spatial metrics. Spatial interactions within the tumor microenvironment are identified as disease-related motifs in a heterocellular system. “If malignant phenotypes supervene upon heterocellular signaling, then perturbing heterocellular nodes and/or edges could present a powerful approach to treat cancer” [46]. With this in mind, we demonstrate how spatial heterogeneity may be used as a cancer biomarker and wielded as a tool for understanding disease progression rather than viewed as a hindrance toward effective treatment. In Chapter 2 we begin by describing a tool for interrogating the TME of high-dimensional tumor images, to quantify ITH and cell-cell

interactions.

1.5 LIST OF PUBLICATIONS

Journal papers

1. D.M. Spagnolo, Y. Al-Kofahi, P. Zhu, T.R. Lezon, A. Gough, A.M. Stern, A.V. Lee, F. Ginty, B. Sarachan, D.L. Taylor, and S.C. Chennubhotla. Platform for quantitative evaluation of spatial intratumoral heterogeneity in multiplexed fluorescence images. *Cancer Res*, 77(21):e71e74, 2017.
2. D.M. Spagnolo, R. Gyanchandani, Y. Al-Kofahi, A.M. Stern, T.R. Lezon, A. Gough, D.E. Meyer, F. Ginty, B. Sarachan, J. Fine, A.V. Lee, D.L. Taylor, and S.C. Chennubhotla. Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. *J Pathol Inform*, 7:47, 2016.

Peer-reviewed conference/workshop papers

1. D.M. Spagnolo, L. Nguyen, J. Bauer, A. Palekar, M. Pejchal, J. Fine, D.L. Taylor, S.C. Chennubhotla. Determining histological tissue image origins with convolutional neural networks International Conference on Machine Learning Workshop on Computational Biology, June 24, 2016, NY.

Manuscripts in progress

1. D.M. Spagnolo, S. Uttam, A.M. Stern, Y. Sui, C. Sevinsky, F. Ginty, D.L. Taylor, and S.C. Chennubhotla, and co-authors. The combination of protein biomarkers and their spatial correlations from hyperplexed IF in colon cancer tissue samples enables predictions of recurrence. 2018.

2.0 VISUALIZATION OF SPATIAL INTRATUMOR HETEROGENEITY

2.1 ABSTRACT

We introduce THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), an open-source tool developed to assist cancer researchers in interactive hypothesis testing. The focus of this tool is to quantify spatial intratumoral heterogeneity (ITH), and the interactions between different cell phenotypes and non-cellular constituents. Specifically, we foresee applications in phenotyping cells within tumor microenvironments, recognizing tumor boundaries, identifying degrees of immune infiltration and epithelial/stromal separation, and identification of heterotypic signaling networks underlying microdomains. The THRIVE platform provides an integrated workflow for analyzing whole slide immunofluorescence images (WSFIs) and tissue microarrays, including algorithms for segmentation, quantification, and heterogeneity analysis. THRIVE promotes flexible deployment, a maintainable code base using open-source libraries, and an extensible framework for customizing algorithms with ease. THRIVE was designed with highly multiplexed immunofluorescence images in mind, and, by providing a platform to efficiently analyze high-dimensional immunofluorescence signals, we hope to advance these data toward mainstream adoption in cancer research.

2.2 INTRODUCTION

Spatial intratumoral heterogeneity (ITH), quantified as the number and variation of cell phenotypes, as well as the spatial relationships between cells and extracellular molecules

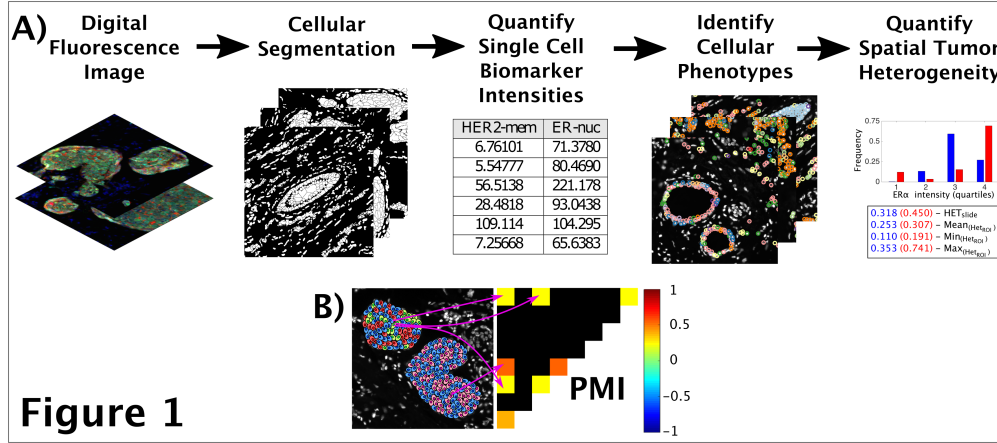


Figure 4: Tumor Heterogeneity Research Interactive Visualization Environment (THRIVE).

(A) For a given panel of images, a cell segmentation algorithm is run to obtain single-cell resolution. Then, biomarker intensity statistics (e.g. mean, median) are computed for each cell from the segmentation results. These statistics are used to discover cell phenotypes via pattern recognition. Heterogeneity metrics are used to quantify the spatial relationships between cell phenotypes. The bar graph shows the heterogeneity of cell phenotypes discovered from ER expression for two different tumor ROIs (shown in red and blue). Phenotype heterogeneity is quantified by quadratic entropy summarized over the whole slide and statistics from ROIs. (B) Pointwise mutual information (PMI) maps capture the relative spatial co-occurrences of cell phenotypes (denoted by various cell colors) in a multiplexed IF image (1). The diagonal elements of the PMI map denote globally heterogeneous and locally homogeneous interactions, while off-diagonal elements capture locally heterogeneous interactions. PMI is scaled from -1 (negative association) to 1 (positive association) where 0 is the background co-occurrence of cell phenotypes.

within a tumor microenvironment (TME), is of high prognostic and diagnostic value [50, 51, 11]. The acknowledgement of spatial ITH as a key factor in tumor progression has identified a need for new informatics tools to quantify spatial heterogeneity in cancer research applications.

Toward this end, we have created an open source tool, THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), which 1) permits visualization of large cohorts of whole slide images and tissue microarrays; 2) performs interactive image analysis tasks such as cell segmentation, cell phenotyping, and tumor microdomain discovery via ITH, and 3) contains statistical inference tools to aid in cancer-specific hypothesis testing. We adopt the term tumor microdomain to describe phenotypically distinct regions of the TME, which represent a fundamental unit of spatial heterogeneity [52]. This software platform encapsulates a workflow for quantifying ITH in immunofluorescence (IF) images ranging from a single biomarker to standard multiplexed biomarkers (up to 7) to emerging hyperplexed (>7) images [30, 53]. Each additional biomarker in IF images allows for more insight into cellular and disease mechanisms, but increases cost and data acquisition complexity, so it was important to develop a platform applicable to a range of imaging modalities. Existing image analysis tools such as CellProfiler [54], ImageJ/Fiji [55], and BioimageXD [56], while useful, are very general tools and thus contain only several of the required features necessary for analyzing spatial ITH, especially from multiplexed and hyperplexed IF images. While some of these contain colocalization pipelines for measuring spatial coincidence of biomarkers within single cells, THRIVE incorporates novel information theoretic measures (pointwise mutual information) and current ecological diversity metrics (quadratic entropy) to enhance insights into the spatial organization of tumors by looking at interactions between cells in the TME [50, 57]. We provide the added benefit of designing algorithms with high dimensional image data in mind, collected through multiplexed immunofluorescence, mass spectrometry, or other data collection methods that allow for a large array of molecular probes. THRIVE allows for the creation of custom workflows with plug-in architecture for new functions, can potentially link to genomic and clinical data, and provides multiple spatial and population based heterogeneity metrics, for ease of use by cancer biologists and clinicians alike.

2.3 THRIVE PLATFORM DESCRIPTION

A computational cancer researcher will find THRIVE to be 1) extensible such that cancer researchers can easily add new experimental tumor heterogeneity algorithms and datasets; 2) maintainable within the research community by leveraging existing open-source libraries, therefore minimizing custom code; and 3) flexible to deploy in a variety of environments, from local research lab installations to cloud deployments shared by the research community. THRIVE uses Docker, which has the advantage of consistent behavior and easy deployment on laptops, computer workstations, and on cloud services.

THRIVE’s file structure is very general, and can easily import files from a variety of microscope platforms. The top-level contains each microscope slide directory; the second level stores directories for each imaged regions in a particular slide, the third level contains both the source image directory and results directories for each region. Under the source image directory are folders for each acquired channel, and under the results directory are folders for each algorithm’s (e.g. segmentation, quantification) output.

Ease of integrating new analysis methods has been an overriding requirement in our design of THRIVE. The straightforward steps that an algorithm developer needs to follow are detailed in THRIVE’s technical documentation. Briefly, the developer would write a short script which pulls input files from data storage, launches the algorithm, stores result, checks for errors, and returns status information. This script is then packaged with the generic THRIVE code in a Docker container. Additionally, the inputs, parameters, outputs, and UI display choices need to be explicitly described, and a Docker Compose file is needed to alert THRIVE that a new algorithm is available. Any programming language can be used to develop heterogeneity algorithms, identified and vetted by the research community, and added to THRIVE. Source code will be available on GitHub, and through the ITCR webpage. THRIVE’s project website is located at <http://ith.csb.pitt.edu>.

2.4 THRIVE CAPABILITIES

THRIVE provides a user interface that enables the researcher to browse and review multi-channel WSFIs, request single cell segmentation and quantification, and review results. The researcher can then run a variety of tumor heterogeneity algorithms, and review and compare those results (Fig. 4A). Our platform enables any number of alternate segmentation, quantification, and heterogeneity algorithms to be integrated into the image processing workflow, both algorithms that we plan to include with THRIVE (e.g. single cell segmentation vs. subcellular-resolution segmentation) and algorithms to be developed and shared by the research community.

THRIVE was developed for the analysis of the spatial distributions of biomarkers, typically using immunofluorescence labeling, and typically in tissue sections on slides. Other methods such as mass spectrometry imaging could also be used to generate compatible images of multiple biomarkers. The following three classes of imaging systems can generate IF images compatible with THRIVE: commercial slide scanning systems, high content screening (HCS) systems, and general purpose microscopy systems. Commercial slide scanning systems with multichannel fluorescence capability include the PerkinElmer Vectra, Leica Aperio FL, Hamamatsu Nanozoomer, and others. In general, these are the fastest and most efficient, as they are optimized for slide scanning. Most HCS systems including the Perkin Elmer Opera, Molecular Devices ImageXpress, ThermoFisher Arrayscan, GE INCell, and others have slide holders and can collect multichannel fluorescence images. HCS systems are also fast and efficient for high volume imaging. General-purpose fluorescence microscope systems from Olympus, Nikon, Zeiss, Leica and others can be used to acquire images from slides, using software packages from the manufacturer, or open-source solutions like Micromanager with the Slide Explorer plug in. These systems are less efficient, but more cost-effective than slide scanners or HCS systems. In all cases, images can be easily saved as TIF files and imported for analysis.

A typical workflow, as demonstrated in [Video 1](#), starts with a cell segmentation step, allowing single-channel and two-channel segmentation of individual cells. If only a cell nuclei channel is available, a single-channel segmentation algorithm delineates the individual nuclei

in the image (e.g. [58]), and then extracts synthetic cell boundary approximations using Voronoi tessellation. When a membrane cell-marker channel is also available, first the cell nuclei are segmented and cell boundary approximations are extracted as described above, and then a watershed algorithm refines the cell boundaries using the additional cell-marker channel data. Each cell is assigned a unique ID, and each sub-cellular pixel is assigned into one of two compartments: nuclear or extra-nuclear. Video 1 can be accessed at <http://movies.aacrjournals.org/video/10.1158/0008-5472.CAN-17-0676/supplementary-video-s1>.

In the biomarker quantification step, cell and subcellular-level statistics (e.g. mean, std-dev, mode, etc.) are computed for each available biomarker, as are cell morphometric features (location, area and cell radius). These measurements can be used to calculate the Pittsburgh Indices [59] and other measures of cell-level phenotypic heterogeneity. Once each cell is described as a single point in a multivariate feature space, phenotypes can be identified through standard clustering techniques. We define phenotypes in this context as the combinations of biomarkers and expression levels in sub-populations of cells. Currently, THRIVE uses basic k-means clustering for cell phenotyping by finding k groups of similar cells in an N dimensional space (where N is the number of biomarkers in the image), and will soon incorporate k-SVD [60] phenotyping used in our previous work [50]. The benefit of k-SVD is that it finds a lower dimensional space with which to group cells into phenotypes, and seeks a sparse representation of the cell data where there is less ambiguity about phenotyping cells that border on two potential phenotypes.

Spatial heterogeneity is characterized by microdomains, which we define as subpopulations of cells clustered together considering not only by the relative populations of the cellular phenotypes within it, but by their spatial distribution. The toolkit contains a starter set of methods to quantify spatial heterogeneity, such as our own technique based on pointwise mutual information (PMI) [50]. Using these methods, users can compare microdomains to one another or to the whole tumor (Fig 4B).

2.5 APPLICATION IN TUMOR HETEROGENEITY

Using the THRIVE cell quantification algorithms, tumor cells, as well as an array of different immune cell types, can be identified. THRIVE can quantify the statistically significant co-occurrences between various cell types within tumor microdomains and at microdomain interfaces which are often associated with known intratumor phenomena. For example, THRIVE can measure the degree to which epithelial and stromal cells are intermixed or spatially separated [49] and can determine the amount of immune infiltration (i.e. the degree to which immune cells invade the TME) within a tumor sample or region of interest (ROI) [61], both of which have prognostic potential. The predictive power of the spatial relationships between various immune cells and tumor cells can be applied as a cancer biomarker for immune infiltration. Additionally, the identification of tumor and non-tumor cells can be used to locate microdomains such that the interfaces between dissimilar microdomains can identify tumor boundaries. This tool will be useful in automating ROI discovery, and assisting pathologists in a computational pathology digital slide workflow.

Notably, THRIVE can be used to identify microdomains containing spatial clusters of network signatures contributed by oncogenic signaling pathways. For example, in the phosphatidylinositol-3-kinase (PI3K) pathway, genetic alterations are found in most invasive breast cancers and PIK3CA mutations are hypothesized to drive carcinogenesis in the breast. Using THRIVE workflows, one can assess the emerging spatial heterogeneity in the PI3K pathway and identify microdomains containing common signatures, e.g. the epithelial-stromal interface PI3K/MAPK signature [62]. Similar efforts to study the MTOR pathway in colorectal cancer [30] could also be assisted by using THRIVE.

We envision that THRIVE will enable the determination of a mechanistic link between spatial intratumoral heterogeneity quantification and cancer progression. It has been shown that neoadjuvant chemotherapy for cancer results in changes in spatial heterogeneity that correlate with poor long-term outcome following adjuvant therapy [11]. Since long-term survival is largely defined by progression to metastatic disease, these results suggest that particular microdomains within the primary tumor impart metastatic potential to a subpopulation of treatment-resistant tumor cells. Implementation of our platform presents a

unique opportunity to identify the heterotypic signaling networks within these metastasis-conferring domains that can lead to robust biomarkers mechanistically linked to disease progression and optimized therapeutic strategies for individual patients.

2.6 CONTRIBUTIONS

The core elements of THRIVE as shown in Fig. 4 and Video 1 are the built-in algorithms for identifying elements of the TME and quantifying spatial ITH, as well as visualizing spatial statistics of high-dimensional bioimages. Toward this end, my contributions to this work are manifold. I led the development of the cell phenotype identification algorithm, and implemented the diversity metrics used for measuring ITH, including quadratic entropy, Shannon entropy, and Simpson’s diversity index. More details on these metrics can be found in Section 3.8.2. Additionally, I implemented many of the visualization tools used by THRIVE, including the diversity heatmap where individual cells were displayed in a color relative to the degree of diversity of their immediate neighborhoods. I also contributed to the original concept of THRIVE, as well as conceiving of the potential applications of the toolbox, which developed through weekly meetings with our U01 collaborators at University of Pittsburgh and GE Global Research. GE Global Research was largely responsible for the software development effort as well as the cell segmentation algorithms.

A crucial next step for THRIVE is to incorporate some of the more sophisticated phenotyping algorithms and diversity metrics I have used to study spatial ITH. Notably, in [50], I argue that pointwise mutual information (PMI) can be particularly helpful in quantifying the spatial cooccurrences of cells within the TME, allowing for network-based measures of ITH. By constructing a cell network for each tumor image, the cell-cell interactions that promote tumor diversity can be visualized and displayed by THRIVE. Additionally the k-SVD phenotyping approach used in [50] may provide better results as image dimensionality scales. In Chapter 3, we explore these methods in greater depth.

2.7 ACKNOWLEDGEMENTS

This work was a collaboration with Yousef Al-Kofahi, Peihong Zhu, Timothy R. Lezon, Albert Gough, Andrew M. Stern, Adrian V. Lee, Fiona Ginty, Brion Sarachan, D. Lansing Taylor and S. Chakra Chennubhotla [63]. In addition to the Department of Computational and Systems Biology at Pitt, collaborators came from the Department of Pharmacology and Chemical Biology, Drug Discovery Institute, and Cancer Institute. At GE Global Research in Niskayuna NY, collaborators came from the Biosciences Organization, as well as the Software Science and Analytics Organization.

S.C. Chennubhotla is supported in part by NIH/NHGRIU54HG008540 and UPMC Center for Commercial Applications of Healthcare Data 711077. D.L. Taylor is supported in part by NIHP30CA047904 and PA DHS4100054875. D.M. Spagnolo is supported in part by NIH NIBIB5T32EB009403-07. The work of D.M. Spagnolo, Y. Al-Kofahi, T.R. Lezon, A. Gough, B. Sarachan, D.L. Taylor, and S.C. Chennubhotla was also supported by grant NIH/NCIU01CA204836.

3.0 QUANTIFICATION OF SPATIAL INTRATUMOR HETEROGENEITY IN BREAST CANCER

3.1 ABSTRACT

Background: Measures of spatial intratumor heterogeneity are potentially important diagnostic biomarkers for cancer progression, proliferation, and response to therapy. Spatial relationships among cells including cancer and stromal cells in the tumor microenvironment (TME) are key contributors to heterogeneity. **Methods:** We demonstrate how to quantify spatial heterogeneity from immunofluorescence pathology samples, using a set of 3 basic breast cancer biomarkers as a test case. We learn a set of dominant biomarker intensity patterns and map the spatial distribution of the biomarker patterns with a network. We then describe the pairwise association statistics for each pattern within the network using point-wise mutual information (PMI) and visually represent heterogeneity with a two-dimensional map. **Results:** We found a salient set of 8 biomarker patterns to describe cellular phenotypes from a tissue microarray cohort containing 4 different breast cancer subtypes. After computing PMI for each pair of biomarker patterns in each patient and tumor replicate, we visualize the interactions that contribute to the resulting association statistics. Then, we demonstrate the potential for using PMI as a diagnostic biomarker, by comparing PMI maps and heterogeneity scores from patients across the 4 different cancer subtypes. Estrogen receptor positive invasive lobular carcinoma patient, AL13-6, exhibited the highest heterogeneity score among those tested, while estrogen receptor negative invasive ductal carcinoma patient, AL13-14, exhibited the lowest heterogeneity score. **Conclusions:** This paper presents an approach for describing intratumor heterogeneity, in a quantitative fashion (via PMI), which departs from the purely qualitative approaches currently used in the clinic.

PMI is generalizable to highly multiplexed/hyperplexed immunofluorescence images, as well as spatial data from complementary in situ methods including FISSEQ and CyTOF, sampling many different components within the TME. We hypothesize that PMI will uncover key spatial interactions in the TME that contribute to disease proliferation and progression.

3.2 INTRODUCTION

For many malignancies, molecular and cellular heterogeneity is a prominent feature among tumors from different patients, between different sites of neoplasia in a single patient and within a single tumor [64]. Intratumor heterogeneity involves phenotypically distinct cancer cell clonal subpopulations and other cell types that include local and bone marrow-derived stromal stem and progenitor cells, subclasses of immune inflammatory cells that are either tumor promoting or tumor-killing, cancer-associated fibroblasts, endothelial cells, and pericytes that comprise the tumor microenvironment (TME) or “tumor tissue system” [59, 27, 65, 7]. The TME can be viewed as an evolving ecosystem where cancer cells engage in heterotypic interactions with these other cell types and use available resources to proliferate and survive [66, 67]. Consistent with this perspective, the spatial relationships among the cell types within the TME (i.e. spatial heterogeneity) appear to be one of the main drivers of disease progression and therapy resistance [51, 11, 68, 69, 70]. Thus, it is imperative to define the spatial heterogeneity within the TME to properly diagnose the specific disease subtype and identify the optimal course of therapy for individual patients.

To date, intratumor heterogeneity has been explored using three major approaches (Fig. 2). The first approach is to take core samples from specific regions of tumors to measure population averages. Heterogeneity is measured by analyzing multiple cores within the tumor. The specific analyses include whole exome sequencing [12, 16, 17, 18], epigenetics [19], proteomics [20, 21], and metabolomics [21] (Fig. 2b). The second approach involves “single cell analyses” using the above methods [22, 23], RNA-Seq [24], imaging [25], or flow cytometry [26] after separation of the cells from the tissue. The third approach uses the spatial resolution of light microscope imaging to maintain spatial context and is coupled

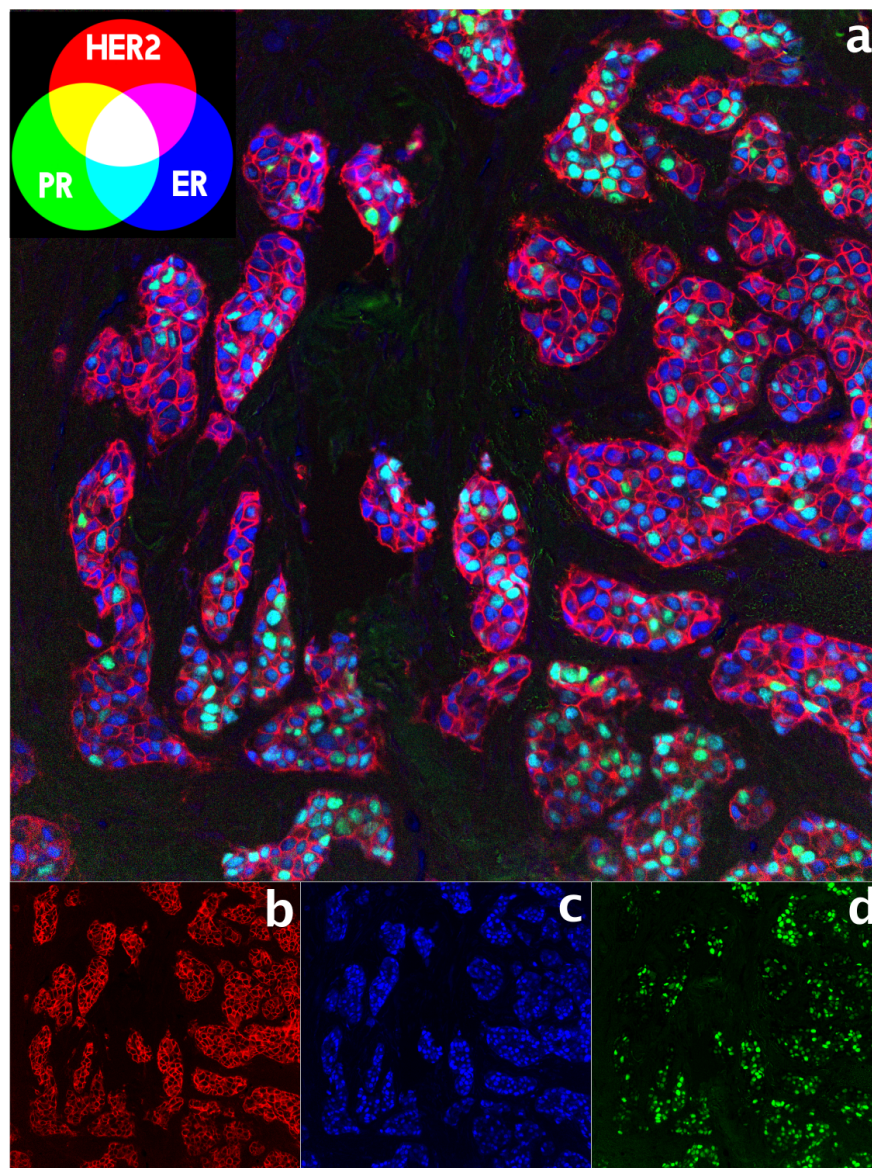


Figure 5: Multiplexed immunofluorescence image of a tissue microarray spot.

(a) Pseudocolored multi-channel fluorescence image is shown for the ER(+) IDC spot 55 of the tissue microarray (Tab. 1). HER2 is shown in red, ER in blue, and PR in green (see legend in the top left corner). Areas of PR/ER co-expression will appear in cyan, HER2/ER co-expression in magenta, and PR/HER2 co-expression in yellow. Arrows point to three different heterogeneous regions in the tumor sample, with varying populations of ER(+), PR(+), and ER(+)/PR(+) cells. The upper arrow indicates a tumor microdomain with higher than average ER(+)/PR(+) phenotyped cells. The middle arrow indicates a microdomain probably best representative of the tumor sample en masse, containing mostly ER(+) cells. In the third microdomain, indicated by the lower arrow, there is a higher than average population of PR(+) cells. (b, c, d) Individual pseudocolored fluorescence images are shown for HER2, ER, and PR.

with molecular-specific labels to measure biomarkers in the cells in situ [27, 28, 29, 71].

Spatial analyses using light microscope imaging facilitate analysis of large areas of tissue sections and/or multiple tumor microarray sections at the cellular and subcellular levels. Subcellular resolution, for example, permits the identification of the activation state of specific biomarkers (e.g., translocation of transcription factors into the nucleus) [53]. In addition, recent developments in mass spectrometry imaging permit many cellular constituents to be measured across a tissue section but at a lower resolution than optical microscopy [72].

Several light microscopy imaging platforms have been developed to characterize cellular biomarker expression levels within tumors including transmitted light and fluorescence [73]. Multivariate information based on fluorescence has been acquired from images of large area tissue sections and tissue microarrays (TMAs) based on DNA, RNA, and protein biomarkers, usually from 1 up to 7 fluorescently labeled biomarkers in the same sample (multiplexed fluorescence) [53, 74]. Multiple commercial platforms can now be used to acquire, process, segment and perform some basic analyses of biomarker signal levels in tissue samples (e.g., PerkinElmer Vectra, Waltham, MA, USA; Genoptix, Carlsbad, CA, USA; Olympus, Center Valley, NJ, USA; Carl Zeiss, Inc., Thornwood, NY, USA; Hamamatsu Photonics, K.K., Hamamatsu City, Japan; Leica Biosystems, Inc., Buffalo Grove, IL, USA). Recently, platforms have been demonstrated permit up to 60 fluorescently labeled antibodies and a few DNA or RNA hybridization probes to be acquired in an iterative cycle of labeling, imaging, and quenching fluorescence [30, 37]. It is now possible to “map” the location of specific cell types, states of cell activations, cell biomarker expression levels, and localizations, as well as extracellular constituents in tissue sections and TMAs.

A major challenge is to develop algorithms that can quantify key spatial relationships (interactions and lack thereof) within the TME based on panels of biomarkers. Initial efforts in measuring heterogeneity in tissue sections applied diversity metrics from ecological studies, such as Shannon entropy and Rao’s quadratic entropy (QE) [11, 75, 76, 77, 78]. However, these methods have not been adapted for multiplexed (up to 7 biomarkers) or hyperplexed (> 7 biomarkers) immunofluorescence (IF) data [59]. Other methods that account for high-dimensional data may not have sophisticated cell phenotyping methods, allowing each biomarker to be only “on” or “off” [79]. Furthermore, a few of these meth-

ods incorporate the spatial relationships between biomarker patterns in their heterogeneity scores [11, 77]. Indeed, the spatial organization of the TME has been hypothesized to be an important diagnostic biomarker in addition to the expression levels of selected biomarkers from both cancer and noncancer cells.

We have designed a method to quantify spatial intratumor heterogeneity (Fig. 2). The method can work with single biomarker, multiplexed, or hyperplexed IF data (Fig. 5). Other heterogeneity characterization methods, although insightful, may not incorporate spatial information or employ multiplexed methods, which implicate a larger number of biomarkers. One such method uses region of interest sampling to add spatial resolution, but this approach does not study spatial relationships between cell phenotypes, nor does it look at more than a single biomarker in its model of heterogeneity [76]. Another method does look at linear relationships among different biomarkers using multiplexed/hyperplexed IF data, but does not incorporate any spatial information, nor does it consider nonlinear associations [32]. Yet another looked at multiplexed phenotypic associations, in contrast to [32] which looked at biomarker associations, but also neglected spatial information [80]. Our method is holistic in its approach, using both the expression and spatial information of an entire tumor tissue section and/or spot in a TMA to characterize spatial associations. In addition, most other methods report intratumor heterogeneity as a single score, thus potentially mapping two spatially different organizations of the TMEs incorrectly to the same score. In comparison, we generate a two-dimensional (2D) heterogeneity map to explicitly elucidate spatial associations of both major and minor subpopulations. We hypothesize that the characterization of spatial intratumor heterogeneity will be an important diagnostic biomarker for cancer progression, proliferation, and response to therapy.

In this paper, the spatial intratumor heterogeneity measure we developed uses data processed on a TMA. We introduce these methods as a proof-of-concept, where we demonstrate the ability to quantify spatial heterogeneity using replicate cores of patient tumors and three breast cancer biomarkers (estrogen receptor [ER], human epidermal growth factor 2 [HER2], and progesterone receptor [PR]) combined with biomarkers for segmentation including the nucleus, plasma membrane, cytoplasm, and epithelial cells (see below). The impact of our method, using pointwise mutual information (PMI) to quantify spatial intratumor hetero-

geneity, will be extended in future studies to the analysis of whole-slide IF images, labeled with increasing numbers of cancer and stromal biomarkers.

3.3 METHODS

3.3.1 Tissue microarray preparation

A TMA of 99 spots (plus orientation cores) consisting of triplicate, 1 mm diameter cores from 24 invasive breast tumor tissues was constructed by the Tissue and Research Pathology Core Facility at the University of Pittsburgh Cancer Institute. Formalin-fixed paraffin-embedded (FFPE) tumor blocks were obtained from the University of Pittsburgh Health Sciences Tissue Bank under the Institutional Review Board approved protocol (PRO13080285). Immunohistochemical expression levels for ER, PR, and HER2 were assessed according to the American Society of Clinical Oncology/College of American Pathologists guidelines. In total, three cases of ER(+) invasive ductal carcinoma (IDC), five cases of ER(+) invasive lobular carcinoma (ILC), eight cases of ER(-) IDC, and eight cases of HER2(+) IDC were each cored in triplicate for a total of 72 tumor tissue spots. The remaining 27 spots consisted of 1 mm cores from cell pellets made from MCF7, MCF10A, MDA-MB-231, and MDA-MB-468 breast cancer cell lines (Breast Cancer Panel, ATCC, Manassas, VA, USA), which were included as staining controls for the multiplexed IF study. Locations of all cores within the array were randomized. As an additional positive staining control, we also purchased a commercial breast cancer tissue array (BRC482 from Pantomics, Inc., Richmond, CA, USA).

3.3.2 Multiplexed immunofluorescence staining and imaging

Multiplexed IF staining and imaging of the study and control microarray slides were conducted as described previously [30]. Two sequential 5 μ m sections of the study array on positively charged glass microscope slides and one BRC482 control slide were taken through the process in parallel. The methods for slide preparation and iterative rounds of staining and imaging were described in detail by Gerdes et al. [30] and issued patents referenced therein.

Briefly, slides were cleared of paraffin, subjected to a 2-step antigen retrieval protocol, and blocked with donkey serum and bovine serum albumin solution. A total of four rounds of staining, imaging, and dye deactivation were completed: Round 1: ribosomal protein S6 (Cy3 conjugate of rabbit anti-phospho-40S ribosomal protein S6 [Ser-240/244], #2215, cell signaling, used at 10 g/mL), ER (Cy5 conjugate of mouse anti-ER-, clone 1D5, #M7047, DAKO, used at 10 g/mL); Round 2: PR (Cy3 conjugate, 13095, used at 5 g/mL), HER2 (Cy5 conjugate of rabbit anti-HER2 clone D8F12.#4290, cell signaling, used at 5 g/mL); Round 3: pan-cytokeratin (pan-CK) (Cy3 conjugate, 13010/13421, 2.5 g/mL), Na⁺/K⁺-ATPase (plasma membrane) (Cy5 conjugate of rabbit anti-sodium-potassium-ATPase, clone EP1845Y, #2047-1, epitomics, used at 5 g/mL); and Round 4: pan-cadherin (PCad) (Cy3 conjugate of rabbit anti-pan cadherin [RB-9036; Thermo Scientific], stained at 5 g/mL), epidermal growth factor receptor (EGFR) (Cy5 conjugate of rabbit-anti-EGFR clone D38B1 [4267; cell signaling], used at 1 g/mL). These eight primary antibodies were conjugated using NHS-ester dye chemistry as previously described, and each lot was revalidated on control tissue sections before used in this study. Background images were collected before the first round and following each dye deactivation step and subsequently used for subtraction of background autofluorescence, as described below. 4',6-diamidino-2-phenylindole (DAPI) staining of nuclei was collected for all rounds of imaging, and DAPI stain was recharged before each antibody staining round. Imaging was on Olympus I 81 inverted fluorescence microscope with a 20 0.75 NA objective outfitted as described by Gerdes et al [30].

3.3.3 Image processing and cell quantification

The biomarker images, acquired in different rounds, were first aligned. Alignment of the different channels was achieved by registering each DAPI image from successive rounds of imaging to the DAPI image of the first round using rigid transformation (i.e., only translation and rotation) [81]. After registration, autofluorescence, which is typical in FFPE tissue, was separated from the fluorophore signals. To do this, an autofluorescence removal process [82] was applied, in which an image of the unstained sample was acquired, normalized, and then subtracted from the corresponding normalized-stained image.

The subsequent step in the workflow was image segmentation, which consisted of several steps (Supplementary Fig. 11). First, DAPI-stained nuclei were segmented using a wavelet-based segmentation algorithm, followed by applying shape-based watershed (Supplementary Fig. 11a,b) [58]. Second, an epithelial segmentation algorithm was used to identify the epithelial cells in the image using a biomarker that is known to be specific to the epithelial cells or at least with known subcellular localization patterns in the epithelial cells (e.g., pan-CK). Third, the cell cytoplasm and membrane were segmented using an algorithm that detects tubular structures in the image based on computing Frangi vesselness (Supplementary Fig. 11c,d) [83]. In parallel, a multi-level watershed algorithm was applied on the membrane segmentation marker to extract initial cell contours [84]. In the final step, initial cell segmentation results were combined with the three individual compartment segmentations (i.e., nuclear, cytoplasm, and membrane) as well as the epithelial mask to generate final cell segmentation mask.

The last step of the workflow was image quantification. Given the segmentation masks and any number of biomarker images, a large number of measurements were computed. These measurements include different cell morphological features (e.g., cell size and shape) and several statistics (e.g., mean, variance, and kurtosis) of each biomarker at the image, cell, and subcellular levels. Although the subcellular measurements for biomarker intensity were available, for this study we used only the mean biomarker intensity at the cellular level. We experimented using the median biomarker intensity as well but observed qualitatively similar results.

3.3.4 Data preparation and automated quality control

After the quantification, two automated quality control (QC) steps were applied. The first QC step detected damaged or lost tissue from round-to-round imaging and artifacts such as light saturation or poor focus, by producing image masks that differentiate between good- and poor-quality regions within each image. The poor-quality regions were subsequently masked out and excluded from downstream analysis. Properly sectioned and imaged cells are expected to fall within a specific size range, and the majority of their nuclei should remain

in the section with minimal fragmentation. Thus, the second QC step filtered out cells based on size and the number of enclosed nuclei fragments. For this work, we used minimum and maximum cell size thresholds of 100 and 3000 pixels (37 and 1110 μm), respectively. In addition, we rejected cells with zero or more than three nuclei fragments [30, 85, 86].

3.3.5 Partitioning cells into two subpopulations to account for the distinct regimes in biomarker intensity distributions

Fig. 6 shows our strategy of quantifying heterogeneity, which is to describe the spatial organization of expression patterns in a tumor sample. To enable this description, we observe that the probability distributions of the mean cellular biomarker intensity values (ER, HER2, PR) can be partitioned into two subpopulations. We will refer the two populations as high- and low-intensity regimes, L1 and L2, respectively (Fig. 7a).

A threshold was defined using the knee observed in the probability distribution plots for each of the biomarkers (shown as vertical lines in Fig. 7a). Since the data had already gone through several stages of QC (as described in the Methods), it was assumed that both the low-intensity and high-intensity populations were unique properties of the data, separate from nonspecific fluorescence. A given cell was assigned to the high-intensity subpopulation if any one of its mean biomarker values was greater than its respective threshold. A cell was assigned to the low-intensity subpopulation if and only if all of its mean biomarker values were below their respective thresholds. Biomarker patterns were then learned for each subpopulation of the data separately.

3.3.6 Learning dominant biomarker intensity pattern set from each cell subpopulation

We aimed to discover a set of biomarker intensity patterns that adequately describe the high- and low-intensity subpopulations (L1 and L2) of IF data. While we could have imposed a predefined set of biomarker intensity patterns (for example, ER high/HER2 low/PR low and ER low/HER2 high/PR low), it was more compelling to learn the dominant patterns from the data, via machine learning (see Methods). The number of patterns needed to describe

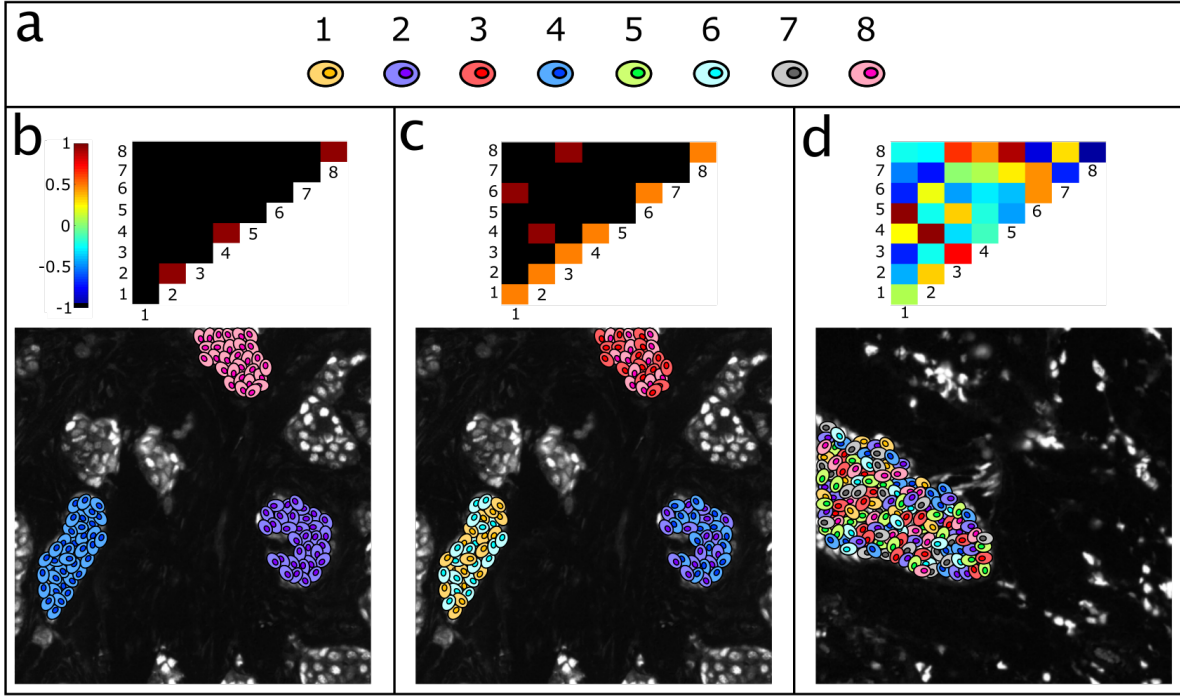


Figure 6: Canonical pointwise mutual information maps depicting various forms of spatial intratumor heterogeneity.

(a) Cartoon representation of eight different cellular phenotypes based on high-dimensional biomarker intensity patterns acquired via pattern recognition algorithms (see Figure 7 for more details). (b) A pointwise mutual information map with strong diagonal entries and weak off-diagonal entries describes a *globally heterogeneous but locally homogeneous* tumor. In this example, the pointwise mutual information map highlights locally homogeneous tumor microdomains containing cells of only one type each, phenotypes 2, 4, and 8 respectively. (c) On the contrary, a pointwise mutual information map with strong off-diagonal entries describes a tumor that is *locally heterogeneous*. In this example, locally heterogeneous tumor microdomains exist as portrayed by the off-diagonal entries. One domain contains phenotypes 1 and 6, another contains phenotypes 2 and 4, and yet another containing phenotypes 3 and 8. (d) Pointwise mutual information maps can also portray anti-associations (e.g., if phenotype 1 never occurs spatially near phenotype 3, see Figures 8 and 9). The ensemble of associations and anti-associations of varying intensities along or off the diagonal represents the true complexity of tumor images in a format that can be summarized and interrogated.

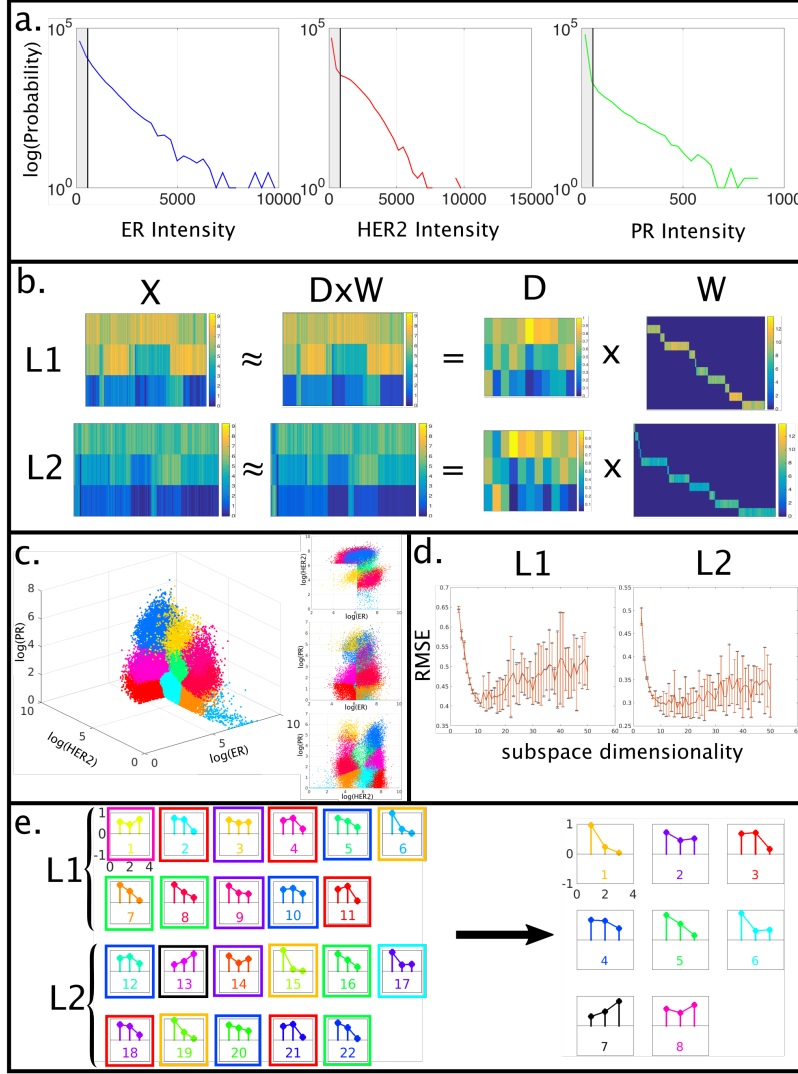


Figure 7: Learning a dictionary of cellular expression patterns.

(a) Thresholds drawn as vertical lines for partitioning dataset into high-signal and low-signal subpopulations (L1 and L2, respectively). (b) Linear approximations of the L1 (high signal) and L2 (low signal) data matrices by the overcomplete dictionaries D and the sparse coding matrices W . Data matrix $D \times W$ is a reconstruction of the dataset, X . The rows of X and D correspond to ER, HER2, and PR biomarker intensities, as labeled. The columns of X and W correspond to each individual cell. The columns of D correspond to the unique dictionary elements and the rows of W correspond to their weights. (c) Each cell is phenotyped to a single pattern in dictionary D . A three-dimensional representation of the L1 matrix is shown, where each cell is color coded by its phenotype. (d) Subspace selection of overcomplete dictionaries D , for L1 and L2, leads to a pattern size of 11 for each subpopulation. (e) Each pattern in the dictionary is shown as a colored stem plot and refers to (from left to right) the ER, HER2, and PR intensity levels. It is convenient to describe these intensities as high, medium, and low as we will do in the main text. For example, the cyan-colored pattern 2 in the L1 dictionary (left), which accounts for the cyan-colored cloud in panel c, may be described as ER high, HER2 high, and PR low. Next, using k-means clustering, we consolidate the L1 and L2 dictionaries into a final dictionary set of size 8. To denote the outcome of k-means clustering, we draw a colored box around each pattern in the L1 and L2 dictionaries, corresponding to the eight different consolidated clusters and show the mean patterns of the consolidated dictionary to the right.

the data is likely to be greater than its dimensionality. This formulation is known as an overcomplete representation (see Supplementary Material for a glossary of machine learning and information theory terms). For example, if ER, HER2, and PR biomarker intensities were binarized into high and low signals, there would be eight potential combinations to describe the biomarkers intensities in three dimensions. In addition, we sought to develop methods that were applicable to single biomarker expression data, multiplexed co-staining methods (< 7 biomarkers), and emerging hyperplexed technologies (> 7 biomarkers). To prescribe a predefined set of patterns on high-dimensional data, one would have to enumerate all potential combinations of biomarker intensities for that dataset: an exponentially complex endeavor. Machine-learning methods have the added benefit of scalability to these higher-dimensional problems.

A previous study suggests using mixtures of Gaussians to model high-dimensional biomarker distributions [80]. However, the formulation we describe below promotes interpretability via the mapping of cells to specific biomarker patterns. In addition, we are making no assumptions about the Gaussianity or distribution of cellular biomarker intensity profiles.

For our pattern recognition, we learned an overcomplete dictionary where the number of biomarker patterns used to represent the data is larger than its dimensionality (number of elements comprising each pattern). For example, we used eight biomarker patterns to describe IF data in three dimensions (ER, HER2, PR). Hereafter, we will refer to an overcomplete dictionary with m patterns as being m -overcomplete. Because the representation was overcomplete, we forced the cells to have a sparse coding, i.e., each cell is phenotyped with only one or few of the biomarker patterns. A sparse coding of the data improves interpretability for both cancer biologists and pathologists.

We used K-SVD, an iterative k-means derived method for designing overcomplete dictionaries with various sparsity constraints, to represent the IF data [60]. Each cell, in the high- or low-intensity subpopulation was approximated by the following linear equation

$$\vec{x}_i \approx w_i^1 \vec{d}_1 + w_i^2 \vec{d}_2 + \dots + w_i^m \vec{d}_m \quad (3.1)$$

with the constraint

$$\text{card}(\vec{w}_i) = 1$$

where $\vec{w}_i = [w_i^1, \dots, w_i^m]$, and is the pattern coefficient vector for the i th cell, $\vec{d}_1, \dots, \vec{d}_m$ represent the patterns in an m -overcomplete dictionary, w_i^j is the pattern coefficient for the j th pattern of the i th cell, and $\text{card}(\vec{w}_i)$ is the cardinality (the number of nonzero elements) in pattern coefficient vector \vec{w}_i . By constraining the cardinality of the pattern coefficient vector to 1, we have explicitly set up the K-SVD algorithm to build a sparse representation such that each cell in a given dataset was phenotyped to only a single pattern in the dictionary (Fig. 7b,c).

The biomarker pattern dictionary was learned using the logarithm of biomarker intensities for each cell in the TMA, for numerical stability. Since K-SVD is an iterative algorithm, we set it to run for 25 iterations well after it had reached a steady state result.

3.3.7 Determine best dictionary size, m

A priori, it is not evident that how many elements should comprise the biomarker pattern dictionary. Representation error of the biomarker intensity data for each cell will tend to decrease as the number of patterns in the vocabulary increases. For example, if the number of patterns is equal to the number of cells in the TMA dataset, the representation error will be zero. However, the representation error of our dataset, given the linear approximations for different vocabulary sizes, decreased at smaller intervals as the vocabulary size approached the number of cells. Thus, we chose a vocabulary size where the representation error was low, but further increases to the vocabulary size had minimal returns. For each potential biomarker pattern dictionary size, m , the error of the linear representation was computed as:

$$e_m = \sum_i \left| \vec{x}_i - \left(w_1 \vec{d}_1 + w_2 \vec{d}_2 + \dots + w_m \vec{d}_m \right) \right| \quad (3.2)$$

To determine the best dictionary size, m , we performed a 10-fold cross-validation on the linear reconstruction for each IF data subpopulation. The data were split into ten equal but distinct groups, and the algorithm was trained on nine parts of the split and tested on the remaining one part. This procedure was repeated to cycle through each of the ten possible testing sets, with both the mean reconstruction error and the variance of the error reported at completion. An elbow criterion was used to choose the best subspace representation,

where the allowance of another pattern in the dictionary was not found to decrease the error significantly (Fig. 7d).

3.3.8 Construct spatial networks to describe the organization of biomarker patterns in a tumor

To represent the spatial organization of the biomarker patterns in a tumor, a network was constructed for each spot in the TMA. The construction of spatial networks for tumor samples intrinsically couples cellular biomarker intensity data (in the nodes of the network) to spatial data (in the edges of the network). The assumptions in the network construction were that cells have the ability to communicate with nearby cells up to a certain limit, up to 250 μm as described by Francis and Palsson [87], and that the ability for cells to communicate within that limit depends on cellular distance. Therefore, the probability distribution was computed for the distance between a cell in the TMA and its ten nearest neighbors. A hard limit was chosen based on the median value of this distribution times 1.5 (to estimate the standard deviation), where cells in the network were connected only within this limit. This limit was consistent with the 250 μm limit proposed by Francis and Palsson [87]. Then, the edges between cells in the network were weighted by the distance between the adjacent cells.

3.3.9 Using pointwise mutual information to quantify spatial biomarker pattern relationships

PMI was used to measure the association between each pair of biomarker patterns in the dictionary and thus different cell phenotypes, for a given sample of the data. This metric captures general statistical association, both linear and nonlinear, where previous studies [32] have used linear metrics such as Spearman’s rho coefficient. PMI may be computed for an individual spot in the TMA, a single patient in the trial (using all of the spots sampled from that patient), a specific cancer cohort in the TMA, or the entire TMA. Once computed for each pair of biomarker patterns, a measure of all associations in the data is displayed in a PMI map. This map describes relationships between different cell phenotypes within

the microenvironment, where differences may be compared from spot-to-spot and patient-to-patient.

Given a linear deconstruction of an IF dataset X , where each column of X is a cell x_k , into an overcomplete dictionary D , where each column of D is a distinct pattern d_i , and a sparse coding matrix W which assigns each cell to only a single biomarker intensity pattern, we assign each cell to have a phenotype f_i , where i is the nonzero index in column w_k of W . A potential pitfall of the algorithm is that high- and low-signal intensity cells can be assigned to the same cell phenotype (more discussion of this in the Results section).

PMI between a pair of biomarker phenotypes (f_i, f_j) for a given network or network set S is defined as:

$$\text{PMI}_s(f_i, f_j) = \log \frac{P(f_{i_s}, f_{j_s})}{P(f_{i_t})P(f_{j_t})} \quad (3.3)$$

where $P(f_{i_s})$ is the probability of phenotype f_i occurring in network set s , and $P(f_{i_t})$ is the background probability distribution of phenotype f_i derived from the complete ensemble of networks. Note that the background distributions are based on the entire dataset, to compare individual networks to the distribution of the TMA as a whole. This construction is similar to the position-specific scoring matrices for either DNA or protein sequences, where the background distributions denote the probability of finding any particular nucleotide or amino acid over the dataset of sequences, for any given position [88]. A PMI map consists of the PMI score for every possible pair of patterns in the vocabulary for a given network set s . While we advocate the interpretation of the 2D PMI map for a thorough understanding of heterogeneity, we also derive a one-dimensional heterogeneity score value from the PMI map, for convenience of the reader interested in comparing with other one-dimensional scores in the literature. The information-deficient one-dimensional heterogeneity score is defined as:

$$\text{HET}_{\text{PMI}_s} = \sum_{i,j} \left| \log \frac{P(f_{i_s}, f_{j_s})}{P(f_{i_t})P(f_{j_t})} \right| \quad (3.4)$$

where higher scores denote a larger difference from the background distribution. The one-dimensional scores can incorrectly map two spatially different organizations of the TMEs, as seen by their PMI maps, to the same scale.

3.3.10 Visualize spatial networks for specific relationships

After computing the PMI map for a given tumor sample or patient and identifying significant interactions or interaction motifs, it is necessary to interrogate the cells which contributed to this significant association. A significant interaction would be considered when the PMI value is close to ± 1 . PMI values close to 1 signify that this particular spatial interaction of biomarker patterns occurs more frequently than observed in the background distribution. PMI values close to -1 signify that when one pattern is observed in the network, the other pattern is found to be observed less frequently than expected from the background distribution. PMI values close to zero signify interactions that may adequately be described by the background distribution.

For a given interaction, a simple linear search through the network can extract the spatial connections in the network that contribute to the associations computed by PMI. These connections may then be superimposed onto the IF image to be examined in more detail by a pathologist (Fig. 8).

3.3.11 Software

We used MATLAB (version R2015a, The MathWorks, Inc., Natick, MA, USA) to implement the analysis pipeline. We applied the segmentation algorithm developed by GE to output the cellular data into a comma separated value file containing the spatial location and the biomarker intensity for each cell in the TMA (Fig. 5) [30, 86]. To partition the data into high- and low-intensity signals (L1 and L2, respectively (Fig. 7a)), we applied a threshold value as determined by the elbow found in the probability distribution of the intensities of each biomarker channel. For biomarker pattern recognition via K-SVD, we used Ron Rubenstein’s MATLAB implementation from <http://www.cs.technion.ac.il/~ronrubin/Software/ksvdbbox13.zip> (Fig. 7b,c). This toolbox also requires the use of an orthogonal matching pursuit implementation in MATLAB from <http://www.cs.technion.ac.il/~ronrubin/Software/ompbox10.zip>. We selected the optimal subspace by performing multiple K-SVD trials at different subspace sizes and chose the ideal subspace dimension when the reconstruction error stopped decreasing (Fig. 7d). To

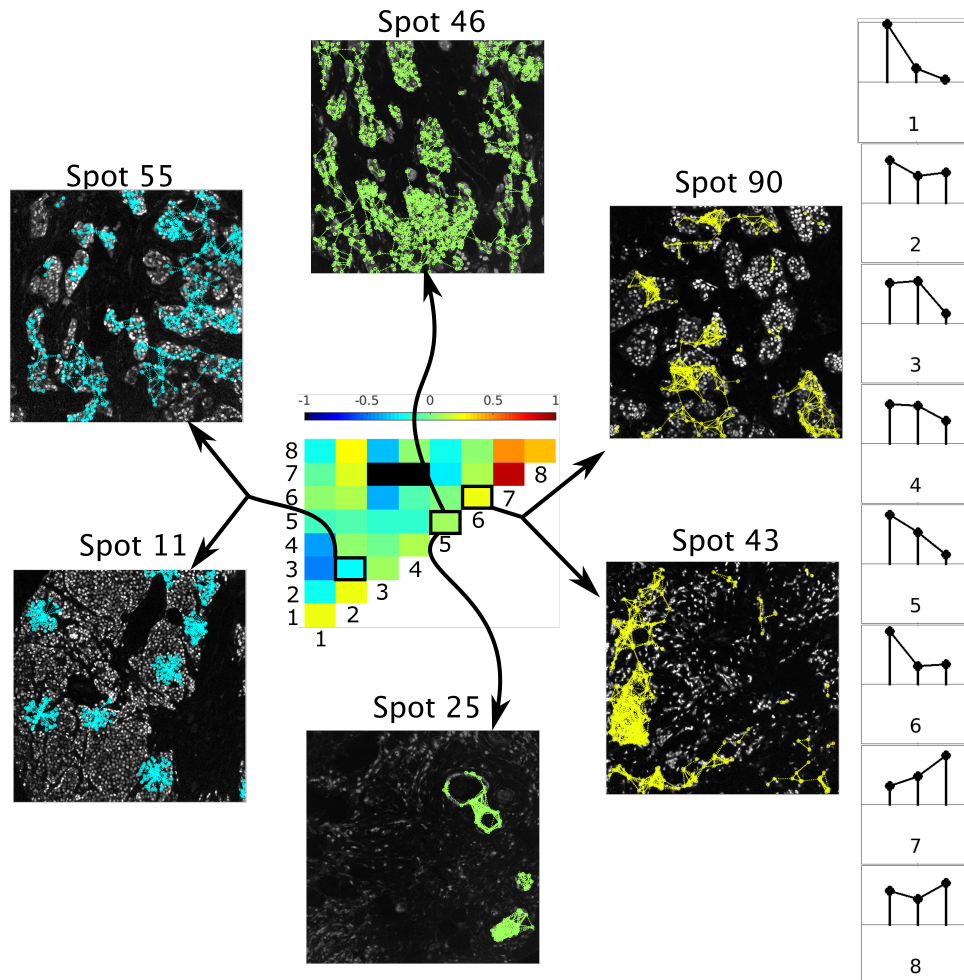


Figure 8: Visualizing networks of spatial interactions from pointwise mutual information maps.

The pointwise mutual information (PMI) map in the middle denotes the relative probability of finding two co-occurring phenotypes i and j in reference to a background distribution. In the colorbar above the PMI map, red/blue indicates highest/lowest possible co-occurrence, and black indicates an absence of interactions. The stem plots to the right describe the eight phenotypes learned from the data, where each stem plot represents the relative ER, HER2, and PR intensities of the phenotype (left to right). The labels for each stem plot (1-8) correspond to the rows and columns of the PMI map. This map allows us to probe any tumor sample for networks of spatial interactions that contribute to the pointwise mutual information calculation. We display representative networks of spatial interactions for three different PMI map entries. The two networks shown in yellow are examples where phenotype 6 spatially co-occurs with itself more frequently than expected from the background distribution. The two networks shown in green indicate two spatial networks where phenotype 5 spatially co-occurs with itself as would be expected from a random phenotyping of cells, given phenotype background probabilities. The two networks shown in blue portray interactions between cells of phenotypes 2 and 3 spatially co-occurring, which happens less than is expected from the background distribution. In each of these cases, the nodes in these graphs are the spatially co-occurring cells of a specific phenotype, and edges are only drawn to cells in spatial proximity. Depending on individual tumor graph statistics, these spatial relationships may be localized or ubiquitous throughout the tumor.

consolidate the biomarker pattern sets learned from the L1 and L2 partitions, we used k-means in MATLAB, where the number of clusters k was chosen using the silhouette function in MATLAB, which provides a graphical representation of cluster membership confidence (Fig. 7e) [89].

To construct cell networks, we computed the median distance, d_m , of the ten nearest neighbors for each cell and derived the standard deviation, s , from the median value, $s = 1.5d_m$ such that the connections between neighboring cells were Gaussian weighted. To keep the network sparse, cells separated by a distance $> 3s$ were not connected. PMI was calculated by Equation 3.3, using 2D histograms to compute the joint probabilities. We observed that the distribution of PMI values was concentrated around 0, ranging between -10 and 10 , with an additional population of extremely large negative values. After it was determined that the extremal values were the result of a lack of co-occurrences in combination with small smoothing coefficients used to calculate numerically stable logarithms, we suppressed these extremal values and focused only on the values around zero. PMI maps were normalized between -1 and 1 for ease of visualization, and the extrema were saturated to a normalized PMI value of -1 . To visualize a specific interaction that contributes to the construction of the PMI map, a linear search may be executed on the tissue data to look for the specific pairwise interactions of cell phenotypes requested (Fig. 8).

3.4 DATASET

Table 1 provides a summary statistics for the TMA with cohorts from (a) ER(+) IDC, (b) ER(+) ILC, (c) ER(-) IDC, and (d) HER2(+) IDC. Three biopsy cores were taken from each patient. For each tumor sample, we consider cells that passed QC (see the Methods section for more details) and divide them into two distinct sets, high and low, based on biomarker intensities.

Table 1: Tissue microarray (TMA) data with cohorts from various breast cancer subtypes.

(a) ER(+) IDC		
Spot #	Patient ID	# Cells
000	AL13-1	23 (1181)
031	AL13-1	222 (382)
026	AL13-2	72 (376)
055	AL13-2	604 (641)
060	AL13-2	420 (624)
005	AL13-3	433 (589)
046	AL13-3	327 (1629)
086	AL13-3	164 (600)

(b) ER(+) ILC		
Spot #	Patient ID	# Cells
001	AL13-4	712 (249)
043	AL13-4	882 (192)
066	AL13-4	1065 (112)
011	AL13-5	2589 (166)
061	AL13-5	3339 (26)
080	AL13-5	2975 (52)
006	AL13-6	297 (621)
025	AL13-6	269 (458)
076	AL13-6	348 (246)
045	AL13-7	260 (192)
030	AL13-8	20 (125)
056	AL13-8	1062 (165)
096	AL13-8	479 (182)

(c) ER(-) IDC		
Spot #	Patient ID	# Cells
002	AL13-9	0 (1032)
039	AL13-9	43 (779)
079	AL13-9	1 (1084)
012	AL13-10	0 (1173)
052	AL13-10	0 (185)
091	AL13-10	0 (1779)
024	AL13-11	0 (849)
049	AL13-11	28 (831)
089	AL13-11	0 (895)
007	AL13-12	3 (381)
034	AL13-12	1 (482)
062	AL13-12	52 (917)
017	AL13-13	17 (84)
036	AL13-13	4 (1055)
072	AL13-13	44 (1219)
020	AL13-14	1 (1322)
044	AL13-14	0 (2296)
071	AL13-14	8 (1414)
032	AL13-15	12 (764)
057	AL13-15	65 (532)
029	AL13-16	118 (876)
067	AL13-16	4 (804)
095	AL13-16	4 (1771)

(d) HER2(+) IDC		
Spot #	Patient ID	# Cells
003	AL13-17	147 (53)
038	AL13-17	209 (36)
068	AL13-17	276 (39)
015	AL13-18	69 (39)
041	AL13-18	771 (184)
078	AL13-18	319 (678)
013	AL13-19	245 (215)
053	AL13-19	824 (123)
090	AL13-19	970 (178)
023	AL13-20	1044 (757)
063	AL13-20	799 (229)
088	AL13-20	669 (101)
008	AL13-21	39 (162)
033	AL13-21	194 (44)
065	AL13-21	97 (7)
018	AL13-22	677 (6)
048	AL13-22	890 (2)
073	AL13-22	521 (5)
021	AL13-23	86 (57)
058	AL13-23	439 (8)
083	AL13-23	1048 (28)
028	AL13-24	126 (1724)
070	AL13-24	64 (1440)
093	AL13-24	309 (2079)

Spot number refers to the location on the TMA. The two numbers reported in the cell count column refers to the number of cells in the high signal and low signal populations, respectively, where the data was partitioned based on biomarker signal intensities (see Figure 7a for more details). (c) Spots 91 and 89 in ER(-) IDC are anomalous in that the ER staining was observed in the normal epithelium, and not in the cancer cells. These spots were erroneously assigned the category of ER(-) IDC by the automated image processing algorithms. These errors could be avoided by having pathologists manually determine regions of interests.

3.5 RESULTS

3.5.1 Preprocessing of multiplexed/hyperplexed immunofluorescence image data

A cellular segmentation algorithm was previously developed and applied to IF data taking advantage of the selectivity of segmentation biomarkers, DAPI, Na^+/K^+ -ATPase, S6, and pan-cadherin [30]. The output of the segmentation algorithm includes cellular masks and subcellular masks for each cell at nuclear, cytoplasmic, and membrane resolution (Supplementary Fig. 11, Fig. 5). Using this segmentation, we extracted biomarker intensity at single cell resolution such that each cell was represented by its spatial coordinates and biomarker intensities, from which interaction networks were built and biomarker intensity patterns were identified.

Fig. 5 shows a pseudocolored IF image where ER signal is colored in blue, HER2 signal in red, and PR signal in green. The pseudocolored image makes colocalizations of biomarker signals within cells explicit and further helps assess heterogeneity. For example, the cytoplasm of a cell will appear either as cyan if ER (blue) colocalizes with PR (green) or as blue if ER (blue) localizes by itself. For spot 55 (shown in Fig. 5), we concluded that a significant degree of heterogeneity is apparent. For demonstration, we highlighted three groups of cells shown by arrows in Fig. 5. In the first group of cells (middle arrow), ER and HER2 were dominant in their respective cellular compartments and did not colocalize. In the second group of cells (top arrow), ER and PR colocalized to produce cyan, with HER2 signal localizing in the membrane. In the final group of cells (bottom arrow), PR and HER2 were dominant in their respective cellular compartments and did not colocalize, with a much smaller portion of ER(+) cells. This visualization example exemplifies a qualitative approach to describing heterogeneity, which was formalized and made quantitative with the PMI measure.

3.5.2 A strategy for quantifying heterogeneity

Our strategy for quantifying heterogeneity has three components (Fig. 6). First, we learn a small set of dominant biomarker intensity patterns, for example, ER high/HER2 high/PR off; from the IF data based on biomarker intensity composition of each cell, we assign it

to one of the dominant patterns. Fig. 6a shows a cartoon representation of possible cell phenotypes. Second, we construct a spatial network to describe the organization of biomarker patterns in a tumor (see Methods). Finally, we quantitate heterogeneity in the form of a PMI map, where the entries measure how frequently a particular spatial interaction between two phenotypes (referenced by the row and column number) occurs in the dataset when compared to the interactions predicted by a random (or background) distribution over all phenotypes. In Fig. 6b-d, PMI entries in red denote a strong spatial association between phenotypes while entries in black denote a lack of any colocalization. PMI entries colored green denote associations that are no better than a random distribution of cell phenotypes over the entire dataset. In addition, PMI maps can portray anti-associations denoted by blue (e.g., if phenotype 1 rarely occurs spatially near phenotype 3) as shown in Fig. 6d.

A PMI map with strong diagonal entries and weak off-diagonal entries describes a globally heterogeneous but locally homogeneous tumor. To illustrate this, we show a canonical PMI map in Fig. 6b where the associations in the diagonal entries for phenotypes 2, 4, and 8 are strong. This implies that these phenotypes are spatially associated with cells of the same phenotype as shown by the composition of the individual microdomains in the tumor sample in Fig. 6b.

On the contrary, a PMI map with strong off-diagonal entries can describe a tumor that is locally heterogeneous. In the canonical PMI map shown in Fig. 6c, the associations between the cellular phenotypes 1 and 6, 2 and 4, and 3 and 8 are spatially localized. In Fig. 6d, we find associations between all phenotypes in the tumor image, and hence, the PMI is thoroughly intermixed. The benefit of PMI maps over existing measures is that the maps evoke a spatial relationship between phenotypes. These provide not only a summary of cellular composition but also an approximation of the tumor topology. For the sake of brevity, we have not included more complicated PMI map examples, but all maps are built off of these simple interactions.

3.5.3 Building a dictionary of biomarker expression patterns

We segregated the data into two partitions based on the distribution of signal intensity for each biomarker, under the assumption that signal intensity indicates true biomarker expression (Fig. 7a). Notice that each of these log-occurrence distributions may be modeled by two or more linear equations. The notch where these two different models would meet is set to be the threshold for that particular biomarker channel and is drawn as vertical lines in the biomarker intensity distribution graphs. For any given cell, if one or more of its biomarker intensities is above threshold, then that cell is classified as Level 1 (L1). If all of the biomarker intensities for any given cell are below the thresholds in their corresponding biomarker channels, then that cell is classified as Level 2 (L2). These two partitions can be interpreted in terms of their signal-to-noise ratio, where L1 has a higher signal-to-noise ratio and L2 has a lower signal-to-noise ratio in comparison. Each partition of cells is used to learn its own set of biomarker patterns. This approach seems particularly judicious given that the distribution of pattern coefficients for L1 and L2 data has different Gaussianity in general (Supplementary Fig. 12). As shown in Fig. 7a, the studied biomarker intensities have long-tailed distributions, so we chose a log-intensity representation to derive a numerically stable pattern recognition algorithm.

For each partition of the data, L1 and L2, we arrived at a sparse signal representation (Fig. 7b). A given data matrix X , where the columns represent each cell in the dataset and the rows represent the log biomarker intensities of each cell (top to bottom, ER, HER2, PR, respectively), can be approximated by the product of matrices D and W . D represents a dictionary of potential biomarker intensity patterns learned from the ensemble of cells in the dataset X , where each column represents one of the patterns learned from the data and each row represents the respective biomarker intensities of each pattern. W is a sparse matrix, which phenotypes each cell in X to a specific pattern in D with a particular scaling coefficient. Thus, each cell (column in W) is represented by only one cell phenotype, which corresponds to the biomarker pattern (column in D) where the sparse code lies. The color spectrum for each matrix varies from blue (low intensity) to yellow (high intensity).

We also display matrix DW to portray the similarity between the actual data matrix

and its reconstruction. By viewing matrices X and DW , which are column sorted by the dictionary element they have the most consensus with, we can observe that each of the biomarker patterns is present in the data. The benefit of this reconstruction of the data is the ability to represent a large array of cell-level data with a small number of interpretable biomarker patterns, describing highly clustered clouds inherent to the dataset as shown in Fig. 7c. Each cell in the three-dimensional log biomarker intensity space is color coded by its phenotype (see caption, Fig. 7c).

The reconstruction error of our linear representation of a given dataset X into dictionary D and dictionary coefficient matrix W highly depends on the dimensionality of D , i.e., the number of patterns that will be used to describe the dataset X . To choose the ideal dimensionality of D , we perform a 10-fold cross-validation of the data reconstruction (Fig. 7d). As is typical in these analyses, we note that as we increase the dimensionality, reconstruction error and the variance of the error decrease until a certain point where the error variance begins to increase with dimensionality. We found that a dictionary size of 11 patterns optimizes both reconstruction error and variance of the error, for both data partitions, L1 and L2.

Having learned a set of 11 patterns for each nonoverlapping partition of the data L1 and L2, we could merge the two dictionaries into a large single dictionary of biomarker intensity patterns that can describe the entire dataset. However, since these patterns were learned separately from partitions deriving from the same dataset, captured under the same experimental conditions, we noted that there were some redundancies between the dictionary learned from L1 data and the dictionary learned from L2 data. Thus, we used k-means clustering to consolidate the large 22-pattern dictionary (with 11 patterns from each partition) into a smaller final dictionary containing only the unique patterns discovered from our approach (Fig. 7e). In Fig. 7e, the 11 patterns learned from L1 and the 11 patterns learned from L2 are shown to the left. Each biomarker pattern is represented as a stem plot of its ER, HER2, and PR intensity, respectively. For convenience, we will describe the intensity patterns in the stem plots as being high, medium, and low. For example, pattern 8 in the L1 dictionary (shown to the left) may be described as ER high, HER2 medium, and PR low.

The outcomes of k-means clustering, shown to the right, result in a final dictionary

dimensionality of 8 biomarker intensity patterns. The final dimensionality was chosen based on the results of a silhouette criterion for clustering evaluation [89]. The boxes around each of the initial patterns to the left signify their cluster membership and correspond to the color of the pattern in the final pattern set on the right. Note that one pattern was unique to partition L2, pattern 7 of the final pattern set, with low ER expression, intermediate HER2 expression, and high PR expression. This demonstrates the value of partitioning the data into two groups, L1 and L2, where patterns dominant in one partition, but not the other, may be elucidated.

3.5.4 Visualizing networks of spatial interactions from pointwise mutual information maps

We generated PMI maps to summarize the relative probabilities of all pairwise spatial interactions within a given tumor sample. In our reconstruction of the cellular IF data, each cell was assigned a specific phenotype by its dominant pattern under sparse coding (see Methods). Each bin of the PMI plot represents the dependence of a cell phenotype upon other phenotypes or itself, relative to the background distribution of the individual biomarkers over the entire dataset. After identifying important spatial dependencies between phenotypes, we can reference the tumor spots and their respective interactions which contribute to the PMI score.

Fig. 8 displays an example PMI map for the entire dataset (Table 1), excluding the tumor samples containing ≥ 100 cells. Each row and column are numbered from 1 to 8, representing one of the 8 potential cell phenotypes displayed to the right (learned via the approach described in Methods and Fig. 7). Each bin of the PMI map is colored from blue to red, according to its PMI score. Scores close to red signify that a particular spatial interaction between two phenotypes (referenced by the row and column number) occurs in the dataset more frequently than a random distribution of all the phenotypes in the data would account for. Scores close to blue signify that a particular spatial interaction between two phenotypes occurs in the dataset much less frequently than a random distribution would account for. If a bin is green, then it signifies that the background distribution of the phenotypes in the

dataset adequately describes the spatial interaction between those two phenotypes. Bins colored in black signify that this potential spatial interaction is not found in this sample of the data.

In Fig. 8, we show networks of spatial interactions that contribute to any one entry in the PMI map. The spatial dependencies shown in green for spots 46 and 25 signify similarity to the background distribution in the co-occurrence of phenotype 5 (ER high, HER2 medium, PR low) with itself. Note that these interactions can be either ubiquitous throughout the tumor sample (spot 46) or localized to specific tumor structures (spot 25). The spatial dependencies shown in light blue for spots 55 and 11 signify lower probability than background distribution in the co-occurrence of phenotype 2 (ER high, HER2 medium, PR medium) with phenotype 3 (ER high, HER2 high, PR low). In comparison, spatial dependencies shown in yellow for spots 90 and 43 signify higher probability than background distribution in the co-occurrence of phenotype 6 (ER high, HER2 low, PR low) with itself. Note that these interactions are spatially localized. Observe that the PMI maps will change when we evaluate the relative probabilities over different subpopulations.

3.5.5 Pointwise mutual information maps as potential diagnostic biomarkers

In Fig. 9, we show the construction of PMI maps for subpopulations of the dataset: patient AL13-3 ER(+) IDC cores (spots) (Fig. 9a), patient AL13-6 ER(+) ILC cores (Fig. 9b), patient AL13-14 ER(-) IDC cores (Fig. 9c), and patient AL13-21 HER2(+) IDC cores (Fig. 9d). For each panel of this figure, we display the PMI maps for the three replicate cores of a given case, then combine the spatial networks for all three cores to build a patient-level PMI map, and finally report scores for each PMI map.

For AL13-3 ER(+) IDC (Fig. 9a), we observe no interactions involving phenotype 7 (ER low, HER2 medium, and PR high) or phenotype 8 (ER medium, HER2 medium, PR high), denoting a lack of ER(-)/PR(+) cells for this patient. There are also a few interactions that have negative PMI scores, denoting a lower likelihood for co-occurrence than predicted by the background distribution (e.g., phenotype 3 [ER high, HER2 high, PR low] with phenotype 4 [ER high, HER2 high, PR medium]). One particular spatial interaction with a

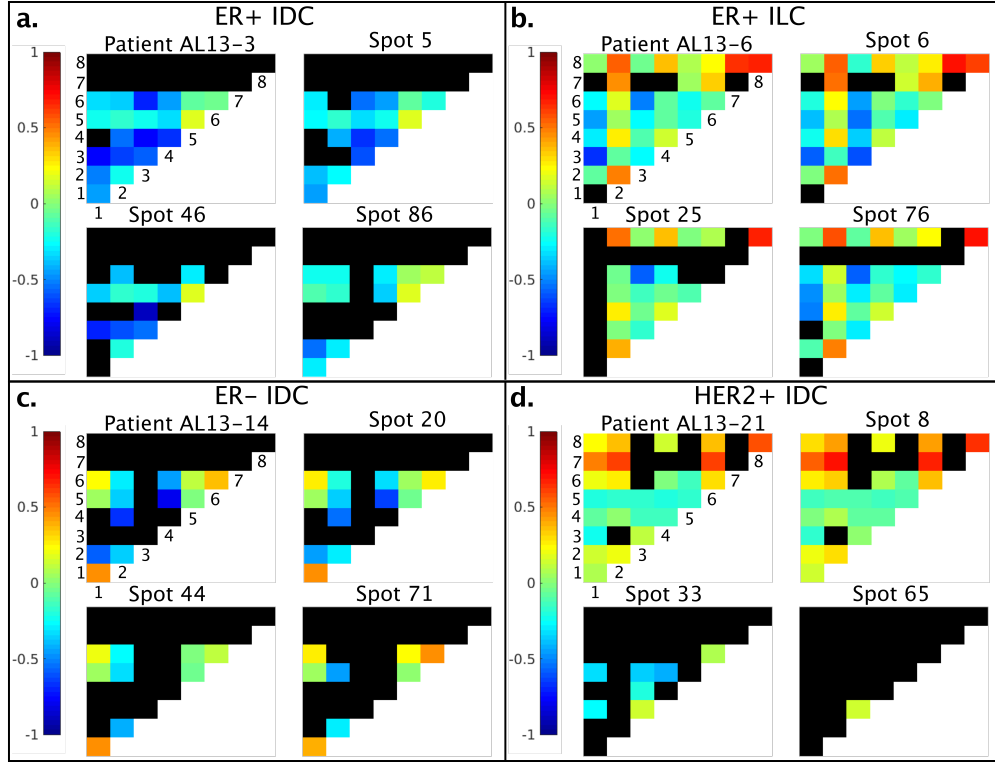


Figure 9: Pointwise mutual information maps as potential diagnostic biomarkers.

PMI maps were constructed for individual cores using the background distributions of cell phenotypes in the entire dataset, and were pooled together for patient-level PMI (entire tumor) to better assess intratumor heterogeneity. A representative (a) ER(+) IDC patient, (b) ER(+) ILC patient, (c) ER(-) IDC patient, and (d) HER2(+) IDC patient PMI map was shown, as well as PMI maps for the three cores taken from each patient. A heterogeneity score was assigned to each core/patient based on the entries in each PMI map (see Methods for the relevant equation). Based on this heterogeneity score, patients AL13-3 ER(+) IDC and AL13-6 ER(+) ILC show more heterogeneity (difference from background distribution) than AL13-14 ER(-) IDC and AL13-21 HER2(+) IDC. The degree to which the core-level PMI maps change with respect to each other and the patient-level map can elucidate how much or little intra-tumor heterogeneity exists. For example, the core-level PMI maps for patient AL13-14 are very similar, signifying that each core is a reasonable approximation for the patient-level analysis. As a contrary example, patient AL13-21 has highly differing core-level PMI maps, signifying a high degree of intra-tumor heterogeneity in this patient. The summary heterogeneity score can provide a simple low-level understanding of heterogeneity between or within patient samples, while the PMI maps can provide a higher-level understanding, providing insight into the spatial relationships of different cell types which brings about the heterogeneity. We also propose visualization tools that can help elucidate these relationships in this higher-level understanding (see Fig. 8).

slightly higher likelihood for co-occurrence compared to the background distribution is that between phenotype 5 (ER high, HER2 medium, PR low) with itself. The observation that the replicate core PMI maps are similar to the combined patient-level PMI map suggests that the cores are reasonably accurate representations of the entire tumor. However, larger PMI values in spot 5 and 46, compared to spot 86, lead to higher heterogeneity scores.

The PMI map for AL13-3 ER(+) ILC (Fig. 9b) presents a high level of deviation from the background distribution as noted by its array of colors. Phenotype 8 (ER medium, HER2 medium, PR high), for example, presents a dynamic behavior, interacting with itself heavily, in addition to phenotype 2 (ER high, HER2 medium, PR medium) and phenotype 7 (ER low, HER2 medium, and PR high), above background distribution. This example suggests that PR(+) phenotypes co-occur spatially. There exists a certain degree of core-level heterogeneity within the tumor. Notably, spots 25 and 76 contain no cells of phenotype 7 (ER low, HER2 medium, and PR high) while spot 6 exhibits phenotype 7 under specific interactions with phenotype 2 (ER high, HER2 medium, PR medium), phenotype 5 (ER high, HER2 medium, PR low), phenotype 6 (ER high, HER2 low, PR low), and phenotype 8 (ER medium, HER2 medium, PR high). Note that, for core 25, the interaction of phenotype 5 (ER high, HER2 medium, PR low) with itself changed from being slightly positive (when the entire dataset was considered, Fig. 8) to slightly negative in the core-level PMI map. Lack of interactions lessens the heterogeneity scores for spots 25 and 76 compared to spot 6. The heterogeneity score for spot 6 is heightened due to stronger anti-associations (for example, between phenotypes 3 and 5 and phenotype 3 with itself), which are then smoothed out by the addition of the two remaining spots for the AL13-6 patient level PMI map.

AL13-14 ER(-) IDC (Fig. 9c) contains fewer PMI interactions than the previous two examples, with 23 of the potential 36 interactions not occurring at all in the patient (shown as black in the PMI map). Phenotype 1 (ER high, HER2 low, PR nearly off) cells, however, co-occur with other phenotype 1 cells more frequently than is described in the background distribution. It may seem counter-intuitive that ER high phenotype 1 occurs in the ER(-) IDC data. However, ER(-) data may very well be captured by an ER high phenotype because a given phenotype decision is made based on the angle between the pattern vectors and the data point in 3D space, and not by the projection distance of the data point onto

the pattern vectors (see Supplementary Fig. 13). The observation that the replicate core PMI maps are similar to the combined patient-level PMI map suggests that the cores are reasonably accurate representations of the entire tumor.

For AL13-21 HER2(+) IDC (Fig. 9d), phenotype 7 (ER low, HER2 medium, and PR high) co-occurs at greater than background probability with phenotype 1 (ER high, HER2 low, PR nearly off), phenotype 2 (ER high, HER2 medium, PR medium), and phenotype 6 (ER high, HER2 low, PR low). In addition, phenotype 8 (ER medium, HER2 medium, PR high) has a highly dependent interaction with itself. An interesting feature of this patient is that the tumor cores taken from this patient are highly heterogeneous. Spot 8 is very dynamic while spot 65 is completely homogenous, containing only phenotype 3 (ER high, HER2 high, PR low) interacting with itself.

Comparing heterogeneity scores across patients, we observe that ER(+) ILC patient AL13-6 has the highest degree of heterogeneity, followed by ER(+) IDC patient AL13-3, HER2(+) IDC patient AL13-21, and finally ER(-) IDC patient AL13-14. Patient AL13-6 has the largest diversity of interactions (few black bins) and contains many strong positive and negative PMI values. Patient AL13-3 has less phenotype co-occurrence diversity but contains mostly strong negative association scores. Following this, AL13-21 contains many interactions with low PMI values. Finally, AL13-14 has the lowest heterogeneity score, containing very few co-occurrences with strong PMI values. Clearly, a breadth of information pertaining to the dependencies between interactions of various cell types and various levels of local and global heterogeneity can be gleaned from core-level PMI maps and comparisons of these maps to their patient-level PMI.

3.6 DISCUSSION

With the ability to capture a growing number of biomarkers, IF and mass spectrometry imaging techniques will play a major role in the quantification of spatial intratumor heterogeneity. These emerging hyperplexed imaging technologies will increase the need for scalable algorithms. The output of these algorithms must remain interpretable and actionable for

decision-making purposes in the diagnostic realm. Our method is flexible regardless of the number of biomarkers imaged and can incorporate spatial and expression data together to quantify spatial intratumor heterogeneity. The end product includes PMI maps, which represent spatial relationships between cell phenotypes and other constituents of the TME. In addition, we provide a score based on PMI values to summarize intra- and inter-tumor heterogeneity. Finally, the PMI maps permit the end user to visualize its individual components in the form of spatially interacting cellular networks.

PMI maps have the potential to be an innovative tool for the modern computational pathologist. For example, if a tumor sample has a PMI map with strong diagonal entries and weak off-diagonal entries, this describes a tumor with several highly localized self-interactions of specific cell phenotypes, thus signifying a tumor sample exhibiting local homogeneity but global heterogeneity. As a contrary example, a tumor sample with a PMI map having strong off-diagonal entries describes a tumor with many localized interactions between different cell phenotypes, signifying a tumor sample exhibiting strong local heterogeneity. Another possibility is the presence of only a single dominant entry in the PMI map, which signifies either global homogeneity if along the diagonal or global heterogeneity with two mixing populations if off-diagonal. By looking at a patient’s PMI map, and comparing it to the individual PMI maps for each tumor sample as well as cohort-summarized (e.g., for all ER(+) IDC patients) PMI maps, unique interactions between biomarkers in the sample can be identified and interrogated. HET_{PMI_s} scores (Equation 3.4) can be used to compare tumor samples quickly. Using these measures, a clinician can arrive at highly specific conclusions regarding the degree of heterogeneity within a single tumor sample or between different samples of the same tumor. The level of information surmised from PMI includes spatial elements that are not achievable by other methods that quantify intratumor heterogeneity, including QE [76].

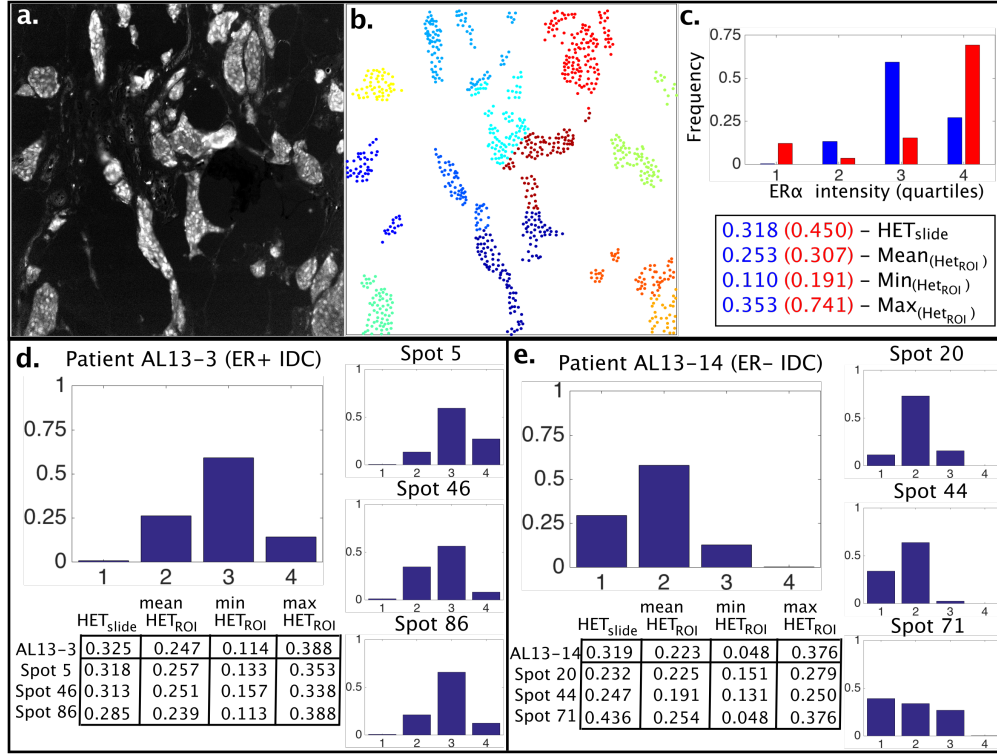


Figure 10: Quadratic entropy implementation with patient and core-level results.

(a) The ER immunofluorescence image is shown for spot 5. Note different levels of ER signal intensity throughout the tumor sample. (b) Regions of interest are selected via k-means clustering on the xy coordinates of each cell (although one may also select regions of interests manually). (c) The frequencies of each ER-intensity species (where 1 is lowest intensity and 4 is highest) are shown for spot 5 by the blue bars. The species frequencies for spot 76 (immunofluorescence and regions of interests not shown) are shown in red to provide another example of potential estrogen receptor-intensity distributions. Quadratic entropy is reported for the entire tumor sample (HET_{slide}), as well as mean, minimum, and maximum quadratic entropy of each regions of interest (HET_{ROI}). (d and e) Quadratic entropy was calculated for patients AL13-3 and AL13-14, from cohorts ER(+) IDC and ER(-) IDC, respectively. HET_{slide} is computed for each individual replicate core of each patient and then computed for the entire patient (consolidating the cells from each core). HET_{ROI} is computed for each region of interest in the cores and the mean, minimum, and maximum HET_{ROI} scores are reported for each core and patient.

3.6.1 Comparing pointwise mutual information to quadratic entropy and clinical standards

QE measures heterogeneity via species diversity (relative abundance) without explicitly quantifying spatial interactions among species. To compare QE with PMI maps, we binned ER signal intensities to define different “species” of cells in the TME. Following work in ecological diversity [75], Potts et al. [76] introduced a distance matrix between these species to penalize spatial interactions between disparate species (more heterogeneity) but reward spatial interactions between similar species (less heterogeneity). High QE values denote more heterogeneity, and low QE values denote less heterogeneity. We computed HET_{slide} as the QE of the entire tumor sample and HET_{ROI} as the QE of various regions of interests in the tumor sample (Fig. 10a-c). We observed for our data that HET_{slide} does not correlate consistently with any of the statistics for HET_{ROI} (Fig. 10d,e). This is similar to our observations in Fig. 9 where patient-level PMI maps do not necessarily correlate to their replicate core PMI maps, highlighting a degree of spatial heterogeneity between cores. Additionally, HET_{slide} can be similar for different tumor samples even though the samples have radically different frequencies of ER expressing species (Fig. 10, e.g., panel d vs. panel e).

One feature of the QE model is the use of a distance matrix between species for characterizing heterogeneity. Another feature of the model is the encoding of spatially heterogeneous subregions in a tumor sample by the computation of both HET_{slide} for the entire tumor sample and HET_{ROI} for the subregions. Finally, the implementation of the QE model is simple and elegant, where species can be defined using only the signal intensity of single biomarkers. However, QE does not quantify the probability of spatial interactions between two cell phenotypes as done in PMI with its spatial network-based approaches. Currently, the QE approach works with one channel at a time while PMI can elucidate spatial relationships simultaneously in the space of all biomarkers. It may also be possible to develop a multichannel QE algorithm to quantify heterogeneity. In addition, QE and PMI are complementary in that a distance matrix can be incorporated into future implementations of PMI. Finally, we intend to incorporate the scaling factor associated with the assignment of cell phenotypes in the construction of PMI maps. It is important to emphasize that PMI maps should be

interpreted in their entirety as 2D maps and not as one-dimensional information-deficient heterogeneity scores.

3.7 CONCLUSION

This paper presents an approach for describing intratumor heterogeneity, in a quantitative fashion, which departs from the purely qualitative approaches currently used in the clinic. With access to larger data samples and clinical outcome data, we will be able to correlate spatial relationships with disease progression and response to therapy. By increasing the number of biomarkers imaged, we can select for cells in different states of activation, as well as noncellular constituents (e.g., secretory elements, extracellular matrix), and quantify relationships between previously unstudied determinants in the TME. In combination with genomic, proteomic, and transcriptomic data, our PMI-based method for spatial intratumor heterogeneity using high-dimensional imaging modalities may be used as part of a multimodal approach to study the mechanisms of cancer.

Having demonstrated that quantification of spatial ITH via PMI can be used as a potential diagnostic biomarker, comparing and contrasting different cancer subtypes (Fig. 9), we now seek the link between ITH and cancer progression, metastatic potential, and response to therapy. Toward this end, in Chapter 4 we explore how spatial ITH measures and other spatial biomarker statistics of high-dimensional tumor images have the ability to predict cancer outcomes. Given patient recurrence and survival data over time, we find a strong concordance between various image feature sets and 5-year recurrence. Additionally, different cancer stages are found to exhibit different spatial statistics. Since metastasis is largely responsible for cancer recurrence and patient mortality, the work presented in Chapter 4 is a step closer toward understanding the mechanisms of cancer metastasis.

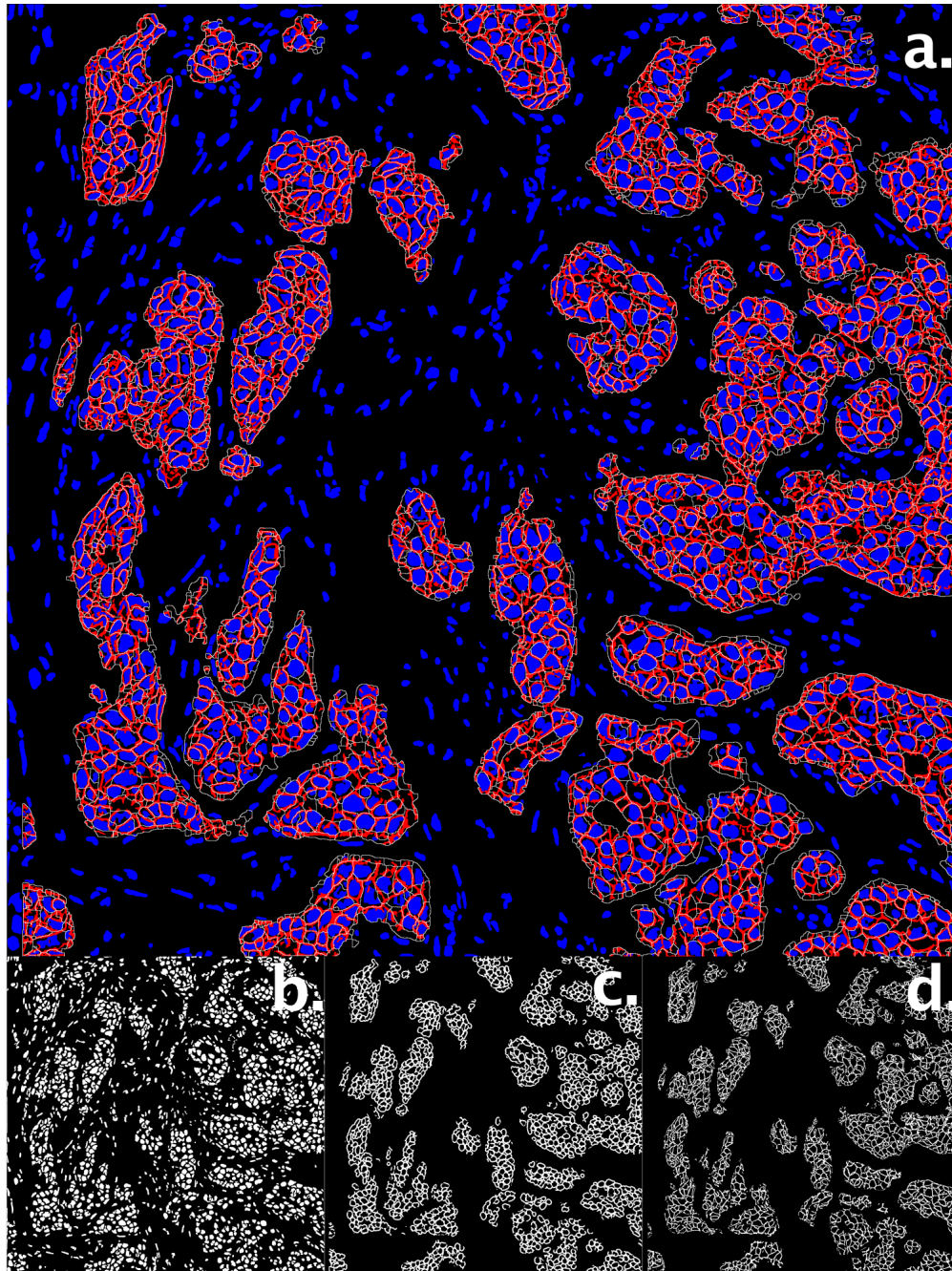


Figure 11: Image segmentation results.

(a) Composite segmentation image is shown for ER(+) IDC spot 55 of the tissue microarray. The nuclear mask is displayed in blue, and the membrane mask is displayed in red. Membrane masks are not shown for stromal cells. (b,c,d) The individual binary masks for nuclear, cytoplasmic, and membrane segmentation are also displayed. The segmentation algorithm is described in Methods.

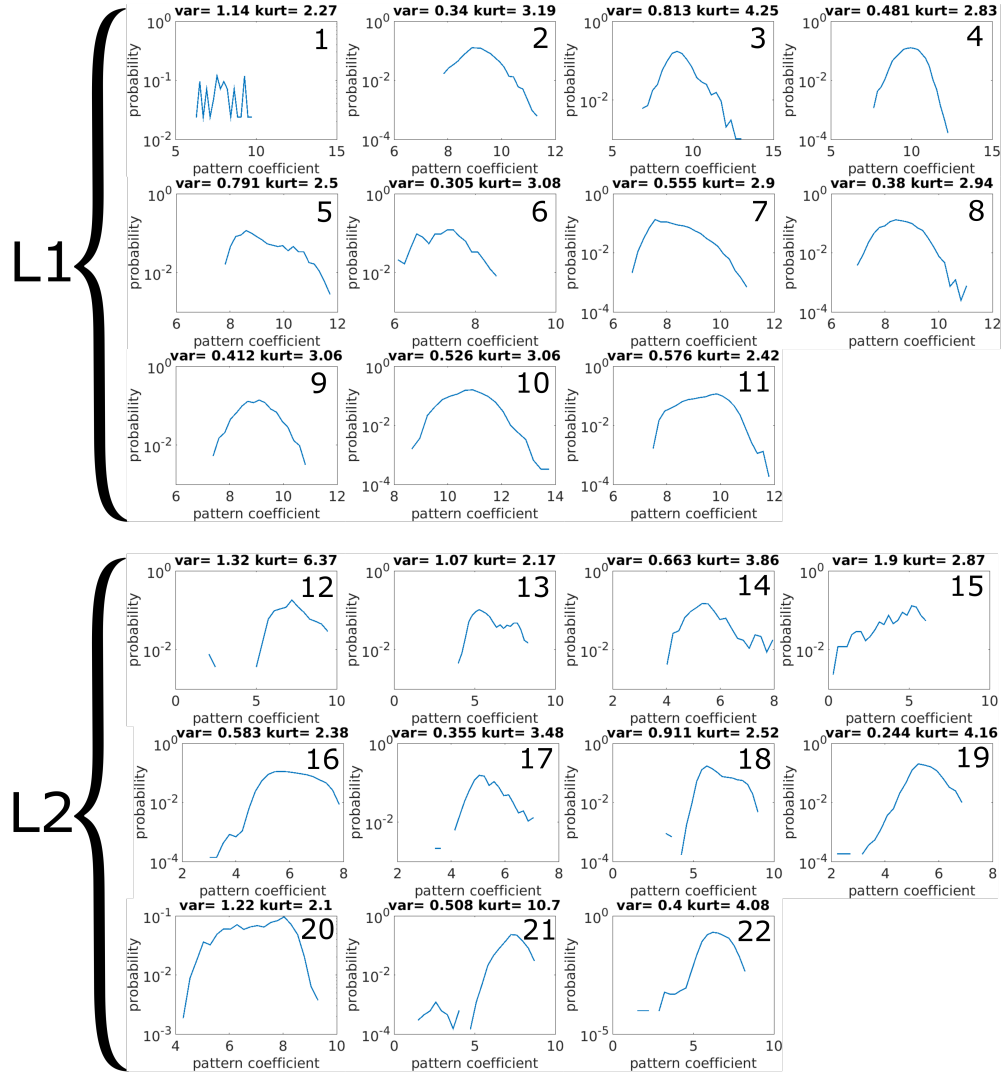


Figure 12: High and low signal subpopulation (L1 and L2) coefficient distributions.

Sparse coding is performed on each partition of the data (L1 and L2) with its respective biomarker pattern dictionary D (patterns 1-11 for L1, patterns 12-22 for L2, see Fig. 7). The distribution of sparse codes (pattern coefficients) is then plotted for each pattern in the L1 and L2 dictionaries, with the coefficient value on the x-axis and log-probability on the y-axis. The variance and kurtosis of each distribution is displayed above its respective plot.

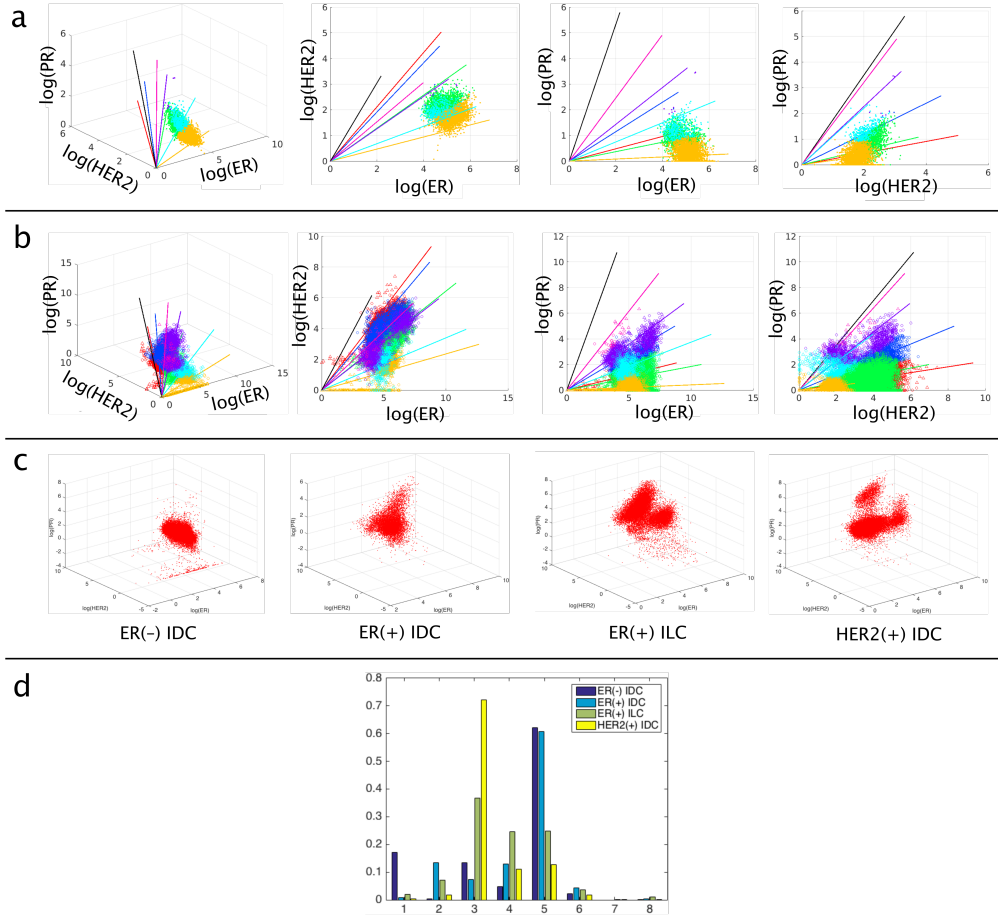


Figure 13: ER(-) IDC distribution in comparison to other cancer cohorts.

(a) ER(-) data plotted in log biomarker intensity space for patient AL13-14 data (spots 20, 44, 71). Note that only patterns 1, 2, 4, 5, and 6 are present in the data cloud: orange, purple, blue, green, and cyan, respectively. There is one blue data point among the purple data cluster. These cells are assigned to the closest biomarker pattern vector based on their angle, but the scaling coefficient to that pattern is typically below a $\log(6)$ value, corresponding to a linear value below 400. (b) 200 points were randomly sampled from ER(-) IDC data (as displayed by unfilled triangles) and 200 points were randomly sampled from ER(+) IDC data (as displayed by unfilled circles). This shows that cells from both data cohorts (ER(-) IDC and ER(+) IDC) can be assigned to similar patterns regardless of their overall biomarker intensity, and dependent only on the angle of their biomarker vector to the biomarker pattern dictionary. (c) It is illustrative to compare ER(-) data with other cancer cohorts. ER(-) IDC shows a dominant cluster that is ER high, HER2 medium, and PR low. Additionally, there is a low intensity cluster that has very low signal value in HER2 and PR. For ER(+) IDC, there is a dominant cluster, similar to ER(-) IDC, as well as a second cluster with larger intensity data. The ER(+) ILC data is organized into two main clusters, separated across a boundary on $\log(\text{HER2}) = 6$, in addition to several lower intensity outliers. HER2(+) IDC is organized into three lobes, one of which contains PR(+) data. (d) The probability of each cell phenotype for the various cancer cohorts. ER(-) IDC is mainly categorized as phenotype 5, with smaller populations of cells in phenotypes 1 and 3. ER(+) IDC is mainly categorized as phenotype 5, with smaller populations of cells in phenotypes 2, 3, and 4. ER(+) ILC is split fairly evenly between phenotypes 3, 4, and 5. Finally HER2(+) IDC is largely categorized as phenotype 3.

3.8 SUPPLEMENTARY MATERIALS

3.8.1 Information Theory & Machine Learning Glossary

Presented in order of appearance:

Rao's Quadratic Entropy (QE) - a measure of ecological diversity (i.e. heterogeneity) for a population divided into a finite number of species. The measure depends both on the relative frequencies of each species, as well as a dissimilarity measure between each pair of species.

Shannon index - originally proposed by information theory founder Claude Shannon to quantify the information content in text strings, the Shannon index has been used to describe the diversity in a population. The more diverse a population is, the more uncertain a prediction of its species frequencies would be. Unlike QE, this index does not account for species dissimilarity.

Elbow criterion - the discontinuity of a plotted function, used to signify that different models are necessary to approximate the function on either side of the discontinuity.

Overcomplete representation - a way to represent real-valued samples with a set of patterns (basis functions) in which the number of patterns is greater than the dimensionality of the input. E.g. our samples are 3D (ER, HER2, PR), and the representation we use to model the structure of the data has a dimensionality of 8.

Overcomplete dictionary - a matrix containing each of the patterns learned in an overcomplete representation of some dataset. Each pattern (column in the matrix) may also be referred to as a dictionary atom.

m-Overcomplete - an overcomplete representation of dimensionality m .

Sparsity - having few non-zero components (in a vector, matrix, etc.).

Sparse coding - a set of unsupervised methods for learning the best representation of a dataset

from a dictionary of patterns. The sparsity of this representation resolves the degenerate nature of an overcomplete representation.

Cardinality - the number of elements in a set. By enforcing a cardinality of 1 for our sparse codes, each cell is assigned only a single cell phenotype.

Orthogonal matching pursuit - an algorithm used to efficiently find a valid sparse coding of a given dataset.

Representation error - the difference (i.e. the information lost) between a dataset and a model of that data.

Subspace - a vector space that is a subset of another higher-dimensional space.

Subspace selection - by observing the change in some metric (e.g. the representation error) through the set of possible subspaces for a dataset, a favorable lower-dimensional subspace may be chosen. Subspace selection is often used in clustering, where the number of clusters needed to represent a dataset is not known a priori and must be chosen.

Cross validation - a model validation technique, which assesses the generalizability of the model. A dataset is split into a training set upon which the model is learned, and a testing set upon which the model is evaluated.

k-fold cross validation - A form of cross validation where the dataset is partitioned into k subsamples. $k-1$ samples are used to train the model, while the remaining 1 sample is used for validation. This process is repeated k times, where each time a different subsample is used for the validation set. A common choice for k is 10, in the case of ten-fold cross validation.

Silhouette analysis - an analysis of clustering, specifically observing how closely samples identify with their own cluster and how loosely they identify with other clusters. The number of clusters, k , used to represent a dataset may be chosen by finding the k which maximizes the average silhouette score.

Pointwise mutual information - a measure of association between two variables, which quan-

tifies the discrepancy between their joint probability and their marginal probabilities. This measure tells us how much our knowledge about the abundance of one variable informs our knowledge about the other.

Spearman's rho coefficient - a nonparametric measure of association between two variables. This measure assumes a monotonic relationship between the variables, an assumption that may or may not hold between different cell phenotypes.

3.8.2 Quadratic Entropy and Other Diversity Measures from Information Theory

Measurements of diversity have been studied extensively in ecology, where they are applied to distributions of individuals or biomass among a finite number of species. These measures, which are rooted in the theory of information, provide us with a starting point for quantifying cellular heterogeneity. A simple method for quantifying diversity is the Shannon entropy, $S = -\sum_i p_i \log p_i$, where the sum is over species and the probabilities sum to one. If we select a random individual from the population, the Shannon entropy measures how uncertain we are about its species. It varies between zero for a homogenous population to $\log N$ when the individuals are equally distributed among N species. The related Simpson index, $\Delta = \sum_i p_i^2$, measures the probability that two randomly selected individuals are of the same species. Hill [90] introduced the idea of “diversity of order q ” which can be expressed in terms of generalized entropies:

$${}^qD = \left(\sum_{i=1}^s p_i^q \right)^{1/1-q}$$

Note that ${}^1D = \exp(S)$ and ${}^2D = \Delta^{-1}$. By using this exponential form, the diversity index provides effective numbers of species [91], enabling easy interpretation. The order of the measure determines the relative weighting of abundant and rare species. For $q = 0$, diversity is essentially a measure of the number of distinct species in the sample. In this limit, rare events are considered important, and loss of rare species has a considerable influence on the quantified diversity. In the limit of q going to infinity, rare species become unimportant

because they lack abundance. The measure in this case is most sensitive to how individuals are distributed among the species with large populations.

By introducing Z as a measure of similarity [57], we can construct a generalized measure of diversity based on the Hill numbers:

$${}^qD^Z = \left(\sum_i i : p_i > 0 p_i (Zp)_i^{q-1} \right)^{1/1-q}$$

The similarity matrix Z makes the diversity index sensitive to the extent of differences between sub-populations; that is, it allows dissimilar species to contribute more to the diversity than similar species. When Z is the identity matrix, the Hill indices above are recovered. It can be seen that replacing $(Zp)_i^{q-1}$ with Dp_i^{1-q} where D is a measure of diversity instead of similarity, we recover the Rao quadratic entropy (QE) for $q = 1$ (see below).

There are differences in biodiversity within an ecosystem and intra-tumor heterogeneity. The ecosystem consists of distinct species whereas the tumor is comprised of a continuum of cellular phenotypes. Further, the distances between species can be measured through phylogeny, but there is no established method for quantifying phenotypic differences within a tumor. Another crucial point is that although biodiversity is well-studied in ecology, less studied is the extent to which species diversity is affected by spatial patterning. The diversity metrics described here were implemented into the THRIVE platform as described in Section 2.4.

3.9 ACKNOWLEDGEMENTS

This work was a collaboration with Rekha Gyanchandani, Yousef Al-Kofahi, Andrew M. Stern, Timothy R. Lezon, Albert Gough, Dan E. Meyer, Fiona Ginty, Brion Sarachan, Jeffrey Fine, Adrian V. Lee, D. Lansing Taylor, and S. Chakra Chennubhotla [50]. In addition to the Department of Computational and Systems Biology at Pitt, collaborators came from the Department of Pharmacology and Chemical Biology, Department of Pathology, Drug

Discovery Institute, and Cancer Institute. At GE Global Research in Niskayuna NY, collaborators came from the Diagnostics, Imaging, and Biomedical Technologies Organization, as well as the Software Science and Analytics Organization.

This work was supported in part by NRSA grant 5T32EB009403-07 (D.M.S), NIH-NCI U01CA204826 (B.S., D.L.T., S.C.C.), PA Department of Health SAP #4100054875 (D.L.T.), NIH Cancer Center-Chemical Biology Facility P30 CA047904 (D.L.T.), Breast Cancer Research Foundation (A.V.L.), internal Pitt-GE grant (A.V.L. and D.L.T.), and U54HG008540 award from NHGRI through BD2K initiative (S.C.C.).

4.0 SPATIAL ANALYSIS OF BIOMARKER EXPRESSION PATTERNS FOR PROGNOSTICATION OF OUTCOME IN COLON CANCER

4.1 ABSTRACT

Colon cancer is a heterogeneous disease, prone to recurrence as a result of metastasis to distant sites. The progression of colon cancer is a result of heterocellular signaling, including the transcriptional reprogramming of the tissue microenvironment into a supportive tumor niche. Notably, many of the previous biomarkers associated with poor outcome have been found to originate in the stroma surrounding malignant cells. We have a unique opportunity to prognosticate outcome in colon cancer with multiplexed immunofluorescence, where we can simultaneously measure a large number of proteins associated with colon cancer signaling up to single cell resolution and map these unique cellular phenotypes spatially. It has been shown that the spatial location of different tissue microarray constituents has different implications for the disease. Using multiplexed immunofluorescence of colon cancer tissue microarray data, we test several multivariate models of protein expression and correlation scores to predict 5 year recurrence in colon cancer. Model building is based on univariate significance of different biomarker features, and prediction is tested using a Cox Proportional Hazards model. We also interrogate which regions of the tumor cores are responsible for high or low risk prediction, potentially highlighting microdomains conferring metastatic potential.

4.2 INTRODUCTION

4.2.1 Colon cancer biology

A 2012 survey named colorectal cancer (CRC) as the third most commonly diagnosed cancer in males and the second in females worldwide, with nearly 700,000 deaths to occur in that year [92]. A more recent US census finds 8% of estimated new cancer cases in men and women to be colorectal, as well as 8% of estimated deaths in the year 2016 [93]. Notably, 35%-45% of colorectal patients with stage II and III cancer succumb to recurrence within 5 years, typically in the form of metastases caused by circulating tumor cells [94].

The heterogeneity of colorectal cancer confounds biological understanding of the disease and results in a diversity of clinical behaviors under the same therapeutic treatment. As a result, oncologists may struggle to decide whether adjuvant chemotherapy is the best course of action for their patient. Notably, colon cancer pathological staging (stage I-IV) does not accurately predict cancer recurrence [95], yet staging is often used as the standard for patient stratification in the case of chemotherapy. After a complete resection of the cancer, stage III patients are often chosen to receive adjuvant therapy (typically FOLFIRI: folinic acid, 5-fluorouracil, irinotecan, or FOLFOX: folinic acid, 5-fluorouracil, oxaliplatin, and previously 5-fluorouracil and levamisole), with increasing evidence supporting adoption by stage II patients [96, 97, 94]. The increasing need for a better standard of care and more effective disease subclassifiers led to a comprehensive molecular characterization effort for cancers of the colon and rectum (similar except anatomically), led in part by The Cancer Genome Atlas [98]. The goal of these works was to have a “fully integrated view of the genetic and genomic changes and their significance for colorectal tumorigenesis.”

From 2012-2015 there were many parallel efforts to subtype colon cancer with unsupervised analysis of exome and transcriptome sequencing, resulting in anywhere from three to six distinct molecular phenotypes. These phenotypes were then post-hoc associated with phenomena including stromal abundance, histology, microsatellite (in)stability, *BRAF* mutation, *KRAS* mutation, epithelial/mesenchymal status, CpG island status, chromosomal (in)stability, stem cell-like behavior, *Wnt* signaling, and DNA mismatch repair [99, 100, 101,

102, 95, 103, 104]. Fortunately, efforts have since been made to reconcile these numerous subtyping schema [105, 106, 107]. An international colorectal cancer subtyping consortium (CRCSC) was formed to interpret 18 CRC data sets (n=4,151), resulting in a four-group consensus molecular subtype (CMS) classification. Although none of the subtypes can be defined by an individual event, and no genetic aberration was limited to a subtype, the taxonomy can be summarized as follows. CMS1 is the MSI immune group, with defining features being microsatellite instability, high CpG island methylator phenotype (CIMP), hypermutation, *BRAF* mutations, immune infiltration and activation, and worse survival after relapse. CMS2 is the canonical group, with high somatic copy number alterations (SCNA), *Wnt* and *Myc* activation. CMS3 is the metabolic group, with mixed MSI status, low SCNA, low CIMP, *KRAS* mutations, and metabolic deregulation. CMS4 is the mesenchymal group, with high SCNA, stromal infiltration, *TGF- β* activation, angiogenesis, and worst relapse-free and overall survival [106].

4.2.2 Tumor microenvironment

Healthy colonic epithelium is “organized into a repetitive crypt structure in which pluripotent intestinal epithelial stem cells either self-renew at the crypt base or differentiate into absorptive enterocytes and goblet cells along the crypt” [46]. Beneath the epithelium lies the lamina propria, containing leukocyte lineages and mesenchymal fibroblasts. In both healthy and malignant colons, epithelial-mesenchymal heterocellular interactions cooperatively achieve tissue-level phenotypes.

With the goal to better understand the “transcriptional reprogramming of stromal cells within tumors,” Scherz-Shouval et al. studied gene programs in cancer-associated fibroblasts (CAFs), the most abundant cells in the tumor microenvironment [108]. In breast cancer, they reported that transcriptional regulator heat shock factor 1 (HSF1) enables malignancy in CAFs, involving stromal signaling molecules *TGF- β* and *SDF1*, where high HSF1 activation associates with poor disease outcome. Calon et al. first demonstrated this effect in CRC in vivo, where mice treated with a pharmacological inhibitor of *TGFBR1* were resilient to metastatic formation [109]. An important finding regarding the role of stroma in CRC

malignance was that gene subsets upregulated in poor-prognosis subtypes [101, 102, 95] “were significantly upregulated in [the] microdissected tumor stroma in comparison to epithelial tumor areas” [110, 111, 112]. CAFs were enriched among the cell types in the poor prognosis gene set, and remarkably there was a linear association between average expression of the CAF gene cluster and risk of relapse after therapy in patients with good-prognosis molecular subtyping. Elevated $TGF-\beta$ was common to all poor-prognosis subtypes, with $TGF-\beta$ in CAFs increasing the tumor initiating capacity of CRC cells [110]. These results reinterpret the role of epithelial-to-mesenchymal transition (EMT) in CRC aggressiveness, but do not necessarily invalidate its occurrence in the epithelium [110, 111, 112].

While stromal elements are essential to regulation or dysregulation of CRC tumors, especially in the development metastatic niches, immune cells also play a critical role to explore further [113]. Galon et al. found the type, density, and location of immune cells within tumor samples to be a better prognosticator of patient survival than CRC staging, with immunologic criteria correlating with survival regardless of tumor extent [114, 115]. Infiltration of CD8+ T cells, referred to as cytotoxic lymphocyte (CTLs), are a factor for good prognosis in CRC, with cancer nest invasion most significantly associating with better survival in comparison to CTL accumulation in the stroma or along the tumoral invasive margin (demonstrating the importance of the spatial organization of the TME) [114, 116]. Another example of the importance of density and location is in tumor associated macrophages (TAMs), with peritumoral TAMs preventing tumor development while intratumoral TAMs resulting in more aggressive cancer cell behavior [117].

The use of multi- to hyper-plexed immunofluorescence will allow quantitative measurement of cellular and molecular signals to accurately portray TME elements and their heterocellular emergence. While previously cellular density and spatial location of these constituents were estimated through microdissection or histology, MxIF will allow for single-cell resolution and accurate characterization of spatial interactions. The importance of accurate signaling localization can be demonstrated via the poor prognosis gene signatures in the consensus molecular subtyping, where assumed epithelial localization led to the hypothesis of EMT but microdissection studies elucidated the stromal origin of these signatures [110]. Another benefit of using MxIF is that morphological information can be harnessed for prog-

nostic potential. The ability to accurately characterize the TME via MxIF provides new opportunities for understanding cancer progression, often through metastasis, and improving patient treatment. “Based on the idea that CRC cells depend on particular stromal factors to effectively disseminate, therapeutics that modulate the CRC ecosystem may be effective in treating or preventing metastasis. An additional advantage of targeting the TME is its genetic stability, making drug resistance less likely to occur” [113]. Understanding the interplay between TME constituents can potentially result in novel therapies, e.g. anti-*TGF- β* therapies and other treatments for disrupting the metastatic niche.

4.3 DATA

4.3.1 Patient statistics

The patients studied in this chapter were collected from Clearview Cancer Institute of Huntsville Alabama from 1993-2002. Of the 747 formalin-fixed paraffin-embedded tumor cores archived (1 core per patient), 694 patients passed rudimentary quality control measures as determined by our collaborators at GE Global Research, and 604 patients passed my additional quality control (including minimum cell number thresholds, and removing patients with non-varying fluorescence signals). The frequency of important clinical variables are displayed in Table 2 as a factor of pathological cancer stage. Among the 604 patients examined in this cohort, the distribution of patients across stages I-III is fairly even (29.5%, 40.7%, and 29.8% respectively). Most patients have grade II colon cancer across the three staging levels. Other variables shown include median age, gender, median days to recurrence (capped at 13 years), median days to survival (capped at 13 years), and number of patients positive for cancer recurrence. The number of lymph nodes positive for cancer cells and the total number of examined nodes were also recorded. Of the 604 patients examined, notably 409 were treated with complete resection alone, with the remaining patients undergoing adjuvant therapy of 5-fluorouracil administration. We are particularly interested in the chemotherapy-naive patient set, as we will be studying the effect of the tumor microen-

vironment on recurrence and survival, without additional intervention upon this niche. The frequency of important clinical variables for the chemotherapy-naïve patient set are found in Table 3.

Pathologists assign each patient to a different cancer stage based on a variety of tests and reports. T stage varies from 1 to 4 and refers to the size of the primary neoplasm and invasion status, with T4 being the largest. N stage refers to the degree of spread to regional lymph nodes, with N0 being no regional lymph node metastasis, N1 representing nodal metastasis at some sites, and N2 representing metastasis to a larger extent. AJCC staging was developed by the American Joint Committee on Cancer, and relies in part on TNM staging (i.e the size of the primary tumor, nodal status, and metastatic spread). AJCC stage I is a CRC that has invaded the muscular layer of the colon, past the mucosa, but has not yet spread to lymph nodes (Stage I = T1 or T2, N0). AJCC stage II has grown to a larger extent, either through the colon wall, rectum wall, or to the abdomen lining (Stage II = T3 or T4, N0). AJCC stage III has spread to the lymph nodes, but has not yet metastasized to distant organs (Stage III = T1-4, N > 0). Cancer grade refers to the morphology of the cells, where G1 refers to healthy “well-differentiated” cells, G2 cells are “moderately differentiated,” and G3 cells are “poorly-differentiated,” looking less like healthy cells [118].

4.3.2 Imaging protocols and quality control

Formalin-fixed paraffin-embedded colon cancer tissue was arranged into three slides of 5 μm sections and imaged using the workflow in Fig. 3, as detailed by Gerdes et al. [30]. The 56 biomarkers imaged with subcellular resolution are listed in Table 6, with functional summaries of each. Images have been registered and autofluorescence removed, and were quantified as described further in subsection 3.3.3. All expression values are \log_2 normalized, and cells with low DAPI signaling are filtered out. Data across the three slides are median normalized, such that the median values of each biomarker across these slides are equal, to account for slide-to-slide staining and imaging variability.

One concern about the imaging protocol of note is that patients from stages I-III were

Table 2: Frequency table for cancer stage vs. cohort variables for complete patient cohort.

Variable	All ($n = 604$)	Stage 1 ($n = 178$)	Stage 2 ($n = 246$)	Stage 3 ($n = 180$)
Age	68(12.3)	66.5(11.9)	70(12)	65(12.5)
Sex	313(51.8%)	90(50.6%)	127(51.6%)	96(53.3%)
Grade 1	87(14.4%)	43(24.2%)	32(13%)	12(6.7%)
Grade 2	444(73.5%)	120(67.4%)	193(78.5%)	131(72.8%)
Grade 3	63(10.4%)	10(5.6%)	17(6.9%)	36(20%)
Recurrence Time	1561.5(1107.2)	1888(1057.5)	1581(1122.1)	1175(1064.4)
Survival Time	1658(1078.2)	1888(1036.7)	1704.5(1115)	1353.5(1024.8)
# of Recurrences	114(18.9%)	13(7.83%)	36(14.6%)	65(36.1%)
# Chemotherapy	195(32.3%)	7(3.9%)	66(26.8%)	122(67.8%)

Age, recurrence in days, survival and days report median value and (standard deviation). Sex, grade, and number of recurrences report total number and (percent). Total number counts for sex are for male patients; female patients would be N - reported male count. Recurrence time and survival time are in days. 1.7% of patients did not have a reported grade.

Table 3: Frequency table for cancer stage vs. cohort variables for all chemotherapy-naive patients.

Variable	All ($n = 409$)	Stage 1 ($n = 171$)	Stage 2 ($n = 180$)	Stage 3 ($n = 58$)
Age	70(12.3)	66(11.8)	73(11.6)	69(13.7)
Sex	207(50.6%)	86(50.3%)	94(52.2%)	27(46.6%)
Grade 1	70(17.1%)	41(24%)	24(13.3%)	5(8.6%)
Grade 2	293(71.6%)	115(67.3%)	141(78.3%)	37(63.8%)
Grade 3	38(9.3%)	10(5.8%)	12(6.7%)	16(27.6%)
Recurrence Time	1636(1139.7)	1899(1062.4)	1562.5(1171)	1046(1080.2)
Survival Time	1668(1136.7)	1899(1051.5)	1651(1186.9)	1146.5(1060.8)
# of Recurrences	57(13.9%)	10(5.8%)	26(14.4%)	21(36.2%)

Age, recurrence in days, survival and days report median value and (standard deviation). Sex, grade, and number of recurrences report total number and (percent). Total number counts for sex are for male patients; female patients would be N - reported male count. Recurrence time and survival time are in days. 2% of patients did not have a reported grade.

not properly randomized across the three slides. As it stands, most stage I patients were placed on slide AGA 260-3, most stage II patients were placed on ATGA 264-3, and most stage III patients were placed on ATGA 269-3. Therefore, a certain degree of speculation must be placed on any result that shows measurable difference between the three stages as this may involve a contribution from slide variability. Median normalization should diminish this effect, but the degree to which it does so is uncertain. Another way we combat the effect of slide variability is by focusing on correlation features over expression features. Protein expression features measure the amplitude of intensity, which may be effected by stain quality and imaging as well, where protein correlation is measuring the angle between two different expression vectors, and should not be effected by staining and imaging variability overall. In this way, correlation features are more robust to intensity changes from potential confounding factors.

4.3.3 Experimental evidence of modeling potential

Before we decided to proceed further with developing multivariate models for predicting CRC recurrence, we needed to assess the modeling potential of our data, including the degree of heterogeneity under various feature sets. In many cases we look at the ability to cluster on cancer staging, as probability of recurrence increases with cancer stage.

The two matrices shown in Fig. 14 represent biomarker expression data for two different tumor cores in the tissue microarray. The columns of each matrix delineate the number of cells in the respective core, while the rows delineate the number of biomarkers. Each cell is represented by a vector of 56 biomarkers cell-level expression values where low expression is displayed as blue and high expression is displayed by yellow. The color bar shows the range of expression values as varying from 0 to 12 (with the maximum log expression being 16 for 16-bit image data). Both epithelial and stromal cells are plotted, epithelial first and stromal second, with the separation point fairly noticeable (near row 1600 in the left matrix, and row 2300 in the right matrix). Both matrices exhibit similar expression vectors across all cells of the same origin (epithelial or stromal), resulting in a row-wise banding effect and representing intra-patient homogeneity. Additionally, both matrices are fairly similar

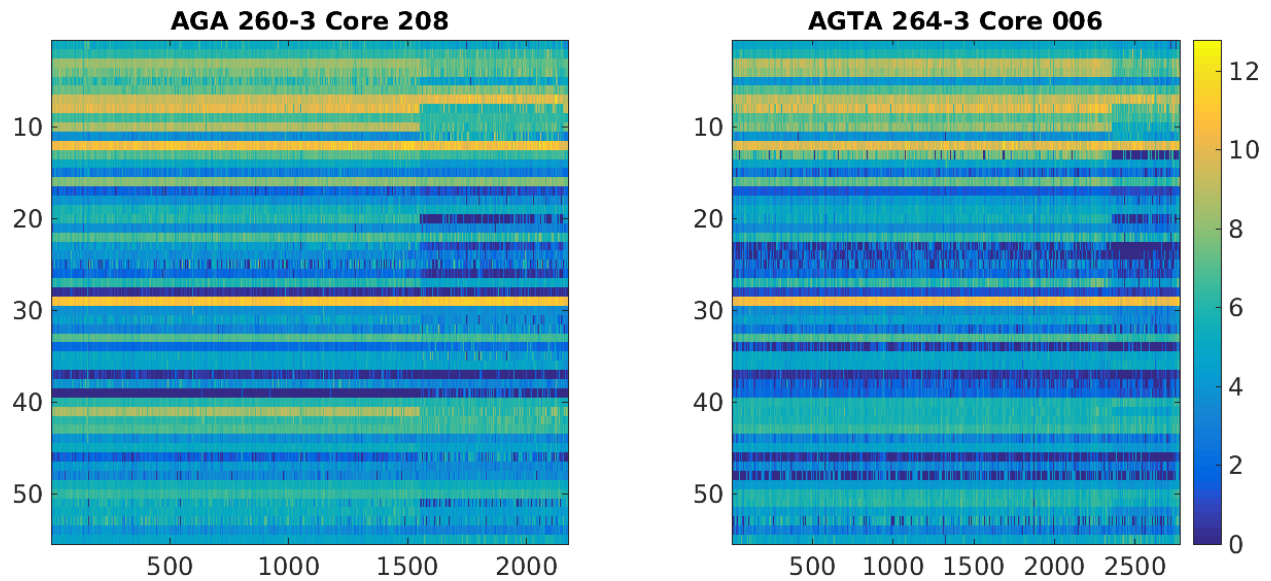


Figure 14: Spatial homogeneity in colon cancer tissue microarray data under biomarker expression features.

On the left is a stage I patient from slide AGA 260-3 core 208, with no recurrence in 5 years. On the right is a stage III patient from slide AGTA 264-3 core 006, undergoing disease recurrence within the first 5 years.

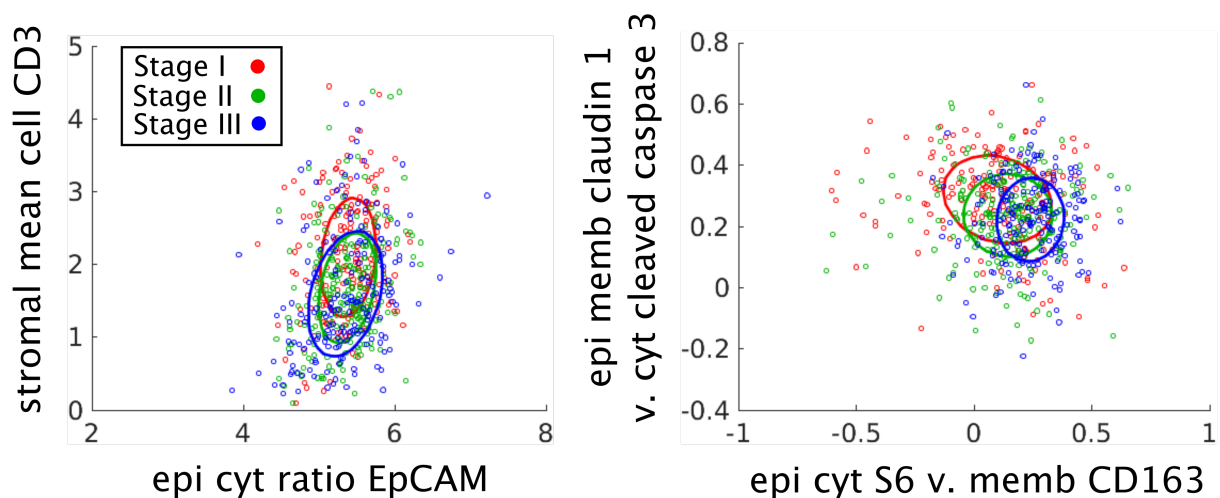


Figure 15: Comparison of scatter plots for expression features vs. correlation features.

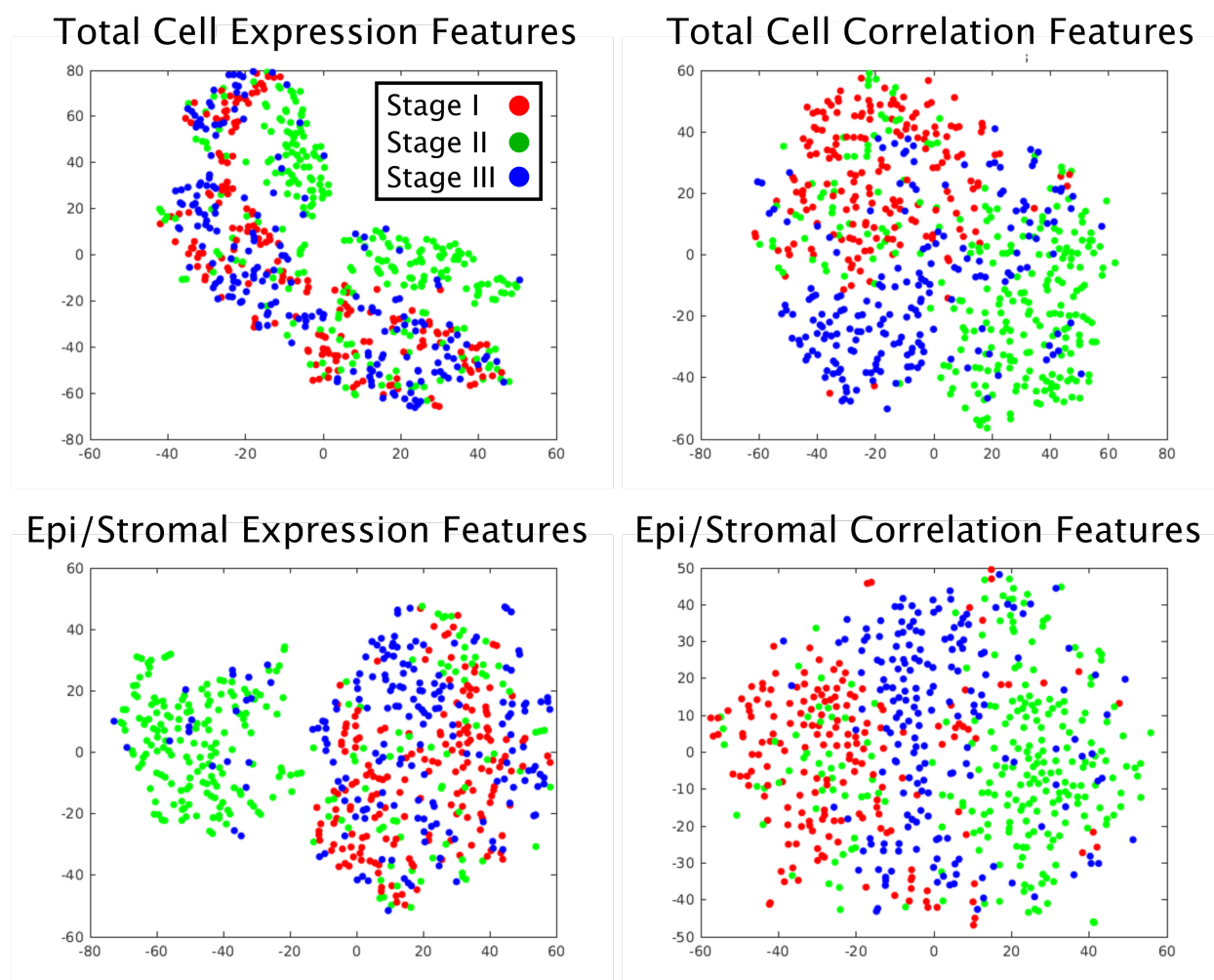


Figure 16: Comparison of t-distributed stochastic neighborhood embedding for expression features vs. correlation features.

to each other, representing inter-patient homogeneity. This is striking, especially since we are comparing a stage I patient with no 5-year recurrence (left) and a stage III patient with 5-year recurrence (right). This suggests that either there is not a significant degree of cellular heterogeneity at the protein expression level, or we may need to explore other modes of representing the potential underlying heterogeneity that leads to differences in disease outcome across patients. The following two figures (Fig. 15 and 16) support the latter hypothesis.

We decided to explore the potential for modeling recurrence using biomarker correlations between pairs of proteins rather than expression values alone. Several benefits of correlations over expression values suggest that this may be a valuable avenue to explore. First, we were concerned about slide-to-slide variability (as discussed earlier), and correlations are more robust to this variability since they measure the angles between expression vectors rather than their intensities. Second, correlation can arguably be interpreted as a spatial feature. With correlation, we are measuring the potential interaction between TME constituents (e.g. CD3 and EpCAM) in the same locale. We can further model the spatial interactions in a tumor core by measuring correlation across a group of cells confined to a certain radius. Third, correlation is a higher order feature compared to expression (a low order feature), and more heterogeneity may exist in this space.

Fig. 15 compares expression features (left) with correlation features (right). For each 2D plot, we are looking at the second and third most significant features for the respective expression and correlation modalities. Each patient is plotted at the mean value of these two features, across all of their cells. Stage I patients are colored in red, stage II in green, and stage III in blue, with a Gaussian density fit to each stage subpopulation. A two-dimensional Kolmogorov-Smirnov test, as described by Fasano & Franceschini [119], was performed against each pair of stages under the expression and correlation features. For the test statistic, values closer to 1 signify rejecting the null hypothesis that the two populations are from the same distribution, and values closer to 0 signify accepting this hypothesis. For expression-based features, stages I and II had a test statistic of 0.2705, stage I vs. III was 0.3408, and stage II vs. III was 0.1586. For correlation-based features, stages I and II had a test statistic of 0.2556, stage I vs. III was 0.3825, and stage II vs. III was 0.2673. Each

of these statistics had a p-value < 0.05 . Thus, for correlation features there was a greater ability to distinguish stages I from III and stage II from III, than with expression features. Note that the scale for expression features is 0-16, while the scale for correlation features is -1 to 1. Also note that the major axis for the three Gaussians under the correlation features are all angled differently, while the three Gaussians for the expression features all fall on the same axis.

Fig. 16 is perhaps most striking in the comparison of expression against correlation features. For each of the 604 patients, a mean expression vector is computed for cell-level features, and a correlation vector is also computed. Then a t-distributed stochastic neighborhood embedding (t-SNE) uses these feature sets to map each sample to a 2D space where similar samples are nearby, and dissimilar samples are distant. For expression features (top left) stage II seems to separate out from stages I and III which are relatively indistinguishable in this space. For the correlation features (top right), all three features are well separated. An interesting feature is that stage I is closer in this embedding to stage III than stage II. However, note that AJCC stage III colon cancers can be T1-4 as long as they have an N stage > 0 , and AJCC stage I cancers are T1 or T2. Thus, AJCC stage I cancers and AJCC stage III cancers can both be T1 and T2, potentially resulting to similarity in this embedding. On the other hand, AJCC stage II cancers must be T3 or T4. It is plausible that stage III is between stages I and II in terms of similarity as stage III can consist of any of the four T stages. When we compute expression and correlation features separately for epithelial and stromal cells (2x the number of features), the banding effect for stages I-III under correlation features is even more noticeable (bottom right). Meanwhile, the same trend persists for expression features, where only stage II separates out (bottom left).

4.4 METHODS

Fig. 17 outlines the steps taken toward the application of biomarker expression and correlation features for predicting recurrence and metastatic propensity. In short, a feature set is generated from the biomarker data, and each feature is tested independently for prog-

nostic power. A salient subset of features is chosen based on this univariate testing, and a subspace selection routine determines the ideal size of a multivariate model from these features. Finally, the multivariate model is used to predict recurrence, and segregate the patient cohort into low-risk and high-risk groups. A more in depth description may be found in the following subsections.

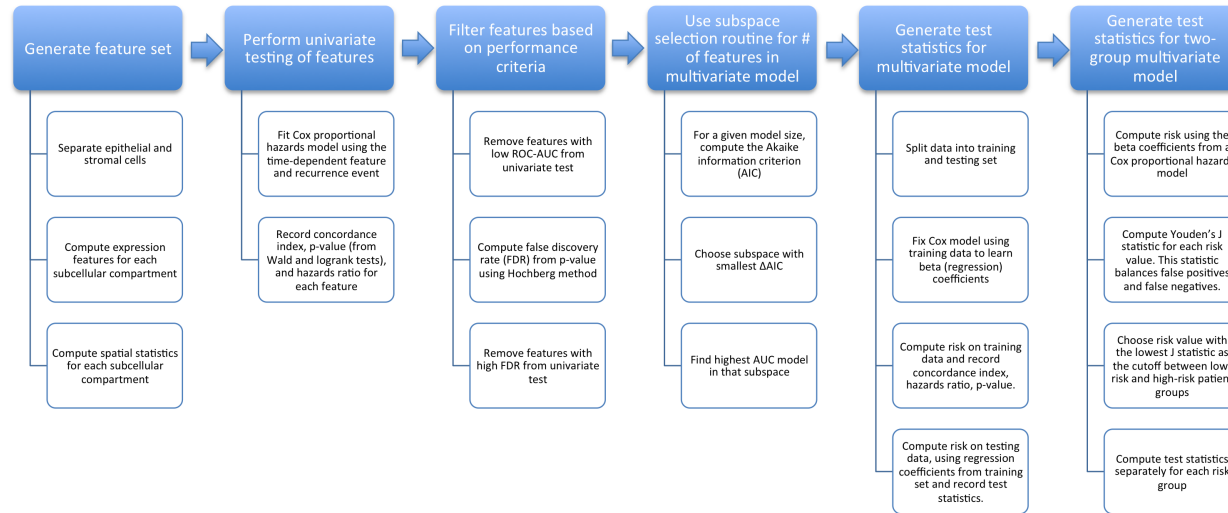


Figure 17: Analysis workflow for recurrence prediction using biomarker features.

4.4.1 Feature sets

Several different feature sets have been tested using the workflow in Fig. 17. Clinicopathology features are the first to consider, because they are the current standard for patient care. This feature set includes patient age, sex, cancer grade (cell morphology score), and AJCC stage (includes tumor extent, nodal spread, and metastatic spread). This feature set will be the baseline against which all others are measured. Next are biomarker expression features, which have gained in popularity in clinic (e.g. Immunoscore [120]). For each patient, we compute the mean expression across all epithelial cells within each of the subcellular compartments. Each biomarker is characterized in terms of mean nuclear intensity, mean cytoplasmic intensity, mean membrane intensity, and mean cell-level intensity. Additionally, nuclear ratio, cytoplasmic ratio, and membrane ratio are also computed, where ratios are computed as mean subcellular location of interest over the average of the other two subcellular means (resulting in 6 epithelial features from each biomarker). Mean stromal expression in the 56 biomarkers are also computed as features, but notably there is only cell-level resolution for these features so only 1 feature is computed here for each biomarker.

Correlation features have been used in localization studies but to our knowledge have not been tested against CRC recurrence, as we have here. For each patient we compute correlation across each pair of the 56 biomarkers in each of the 3 subcellular compartments (nucleus, cytoplasm, membrane) using the epithelial cell set. Correlations are computed separately for stromal cells resulting in a 56 x 56 correlation matrix. Notably we tested three different combination of correlation features under the following two steps of the analysis workflow: univariate testing, feature filtering, and subspace selection (Fig. 17). First we tried exclusively using cell-level features (ignoring subcellular localization). This resulted in suboptimal predictive results. Then we tried using all subcellular localizations for our epithelial correlation matrix (168 x 168). This performed well, but inhibited multivariate model development as exhaustive subspace tests were computationally untenable, and often multiple, likely redundant, features from the same marker ranked among the best features. Finally a compromise was reached by picking one subcellular feature from each biomarker based upon expected subcellular localization (gleaned from GeneCards, PhosphositePlus,

and assorted literature). Thus both the epithelial and stromal correlation matrices were 56x56 but with more specific information retained from our subcellular localization. This is the correlation feature set used going forward. Note, we are using Pearson correlation, but rank correlation or Spearman’s Rho can also be used, if one wanted to capture nonlinear dependencies.

4.4.2 Univariate modeling

Of the 56 biomarkers imaged, we decided to drop our DAPI signal since this is strictly a segmentation marker, with no known significance in CRC. The remaining 55 biomarkers, listed in Table 6 result in 2970 correlation features (55 x 55 subtracting the diagonal) for each cell set, leading to 5940 correlation features overall (2970 epithelial features and 2970 stromal features). For the biomarker expression features, there are 330 epithelial features (6 features for each of the 55 markers) and 55 stromal features, 385 total.

Each feature was individually tested for its ability to predict 5-year recurrence with a univariate Cox Proportional Hazards model. A Cox Proportional Hazards model tests how one or several covariates influences the rate of a particular event (e.g. survival, recurrence). This model is linear for log-hazards (or multiplicative for hazard) with an unspecified baseline hazard, and is defined as:

$$h_i(t) = h_0(t) \exp \left(\beta_1 x_{i1} + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \right) \quad (4.1)$$

where $h_0(t)$ is the baseline hazard, and x_1, \dots, x_k are the covariates. For two observations i and j the hazards ratio is

$$\frac{h_i(t)}{h_j(t)} = \frac{\exp \left(\beta_1 x_{i1} + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \right)}{\exp \left(\beta_1 x_{j1} + \beta_1 x_{j1} + \dots + \beta_k x_{jk} \right)} \quad (4.2)$$

. For each univariate model, we record concordance index, and p-value. Concordance index (C-ind) measures the concordance between the model prediction and the observed event. For each pair of patients with opposing recurrence events, concordance is added if the predicted risk is higher for the recurrent patient over the non-recurring patient [121]. The p-value is computed from a Wald Test, which tests how likely the beta coefficients in the hazards ratio

are different from zero, thus contributing to the prediction of hazard. We also record the p-value under a logrank test, although the two tests are asymptotically equivalent.

4.4.3 Feature selection

Of the 5940 correlation features and 385 expression features, we need to discover which features perform well for potential inclusion in a multivariate model.

Features are filtered out based on two performance criteria, concordance index and false discovery rate. In hypothesis testing, the false discovery rate is the rate of significant predictions that are actually null, i.e. it is the expected value of the ratio of false positives to the total number of predictions (false positives and true positives). The Benjamin Hochberg method was used to correct the obtained p-values, decreasing the false discovery rate (q-values). Features were filtered out (removed) from downstream analysis if they had a concordance index lower than 0.6 and a q-value greater than 0.01.

After arriving at a smaller subset of predictive feature, we must determine how many different features are ideal for a multivariate Cox model for predicting 5-year recurrence in CRC. With too many features one can overfit the data, but with too few features performance can potentially be improved. Starting from single feature models and then increasing in dimensionality, all features in this filtered subset are tested exhaustively. For each feature (or combination of features) tested, concordance index is recorded, as is the Akaike information criterion (AIC). The AIC measures the quality of a statistical model by its goodness of fit (from log-likelihood) and simplicity (by number of parameters). Then, Δ AIC scores between subspace sizes are used to decide the stop point for a salient model, in terms of # dimensions. For our correlations features, we found a 7 feature model to be appropriate, and chose the best 7 feature model using the concordance index.

4.4.4 Multivariate modeling

To properly evaluate our multivariate model, we must split the data into a training set (using 70% of the chemotherapy-naive patient cohort), and a testing set (using the remaining 30%). It is important to use the chemotherapy-naive patients since adjuvant therapy will have a

potential effect on recurrence beyond the population statistics and spatial interactions of the tissue microenvironment. Once the data is split into a training and testing set, we learn the beta coefficients of the Cox model, i.e. the contribution of each covariate toward the risk. Then we use those learned coefficients to test the predictive power of the covariates in the test set for assessing patient recurrence. For both the training and testing set, we record concordance index, hazards ration, and p-values under the Wald and logrank tests.

4.4.5 2-level cox model

Additionally, we can use the risk from our Cox models to partition our patient cohort into low-risk and high-risk subsets, assessing the predictive power of our models for each subset [122]. First, risk is computed for each chemotherapy-naive patient. Then Youden’s J statistic is computed for each patient based on risk. Youden’s J balances the sensitivity and specificity based on the receiver operating characteristic. The risk value associated with the lowest J statistic is then used as a cutoff between low and high risk patients. For each risk subset, we can compute concordance and significance values.

4.5 RESULTS

4.5.1 Comparison of 11 models

After the model building workflow (as described in Fig. 17), we report results from nine different prognostic models.

1. Clinicopathology features
2. Correlation model (7 biomarker pairs)
3. Correlation model (7 biomarker pairs) + clinical
4. Enhanced correlation model
5. Enhanced correlation model + clinical
6. Biomarker intensities (best 7 features)
7. Biomarker intensities (best 7 features) + clinical

8. Biomarker intensities (best 12 features)
9. Biomarker intensities (best 12 features) + clinical

Model 1 includes the patient features of age, sex, AJCC stage, and cancer grade. Model 2 was optimized based on the colon workflow, specifically in terms of the number of features (chosen via ΔAIC) and the optimal model (chosen via exhaustively testing C-ind for all 7 feature models). Model 3 contains all of the model 2 features plus the model 1 clinicopathology features. Skipping ahead to models 6 and 7, we are testing expression features, using a comparable number of expression features to the correlation features in model 2. Model 6 is just using the 7 best expression features from univariate testing, where model 7 also includes clinicopathology features. Models 8 and 9 are to be compared to the best correlation model based on the number of unique biomarkers (12) implicated in the 7 feature correlation model. Thus model 8 uses the 12 best expression features, and model 9 also includes clinical features. We also tested an expression based model using the exact 12 biomarkers implicated in the correlation model but this model did not perform well enough to report. Expression features and the correlation features that may implicate the same biomarkers as cofactors are not necessarily both significant.

The remaining models, 4 and 5 were an experiment. Instead of using the 55 x 55 correlation matrix, where each biomarker was only quantified in its expected subcellular compartment, we used the full 165 x 165 correlation matrix and handpicked features in the following way. We chose the best 7 features model based on concordance index, and then we added biomarker features based on plots of the p-value and concordance index as one more feature was added. We added each feature that resulted in a significant jump in the evaluative score. Although this isn't an automated model building schema, if it performs better, it demonstrates that our automated models can improve even further. Model 4 contains these handpicked correlation features, and model 5 also includes clinical features.

Concordance index, coefficient of determination R^2 , hazards ratio, and confidence intervals are recorded for each of the nine prognostic models in Tables 4 and 5. The gray row signifies a model using only clinicopathology features, while red rows signify models containing correlation-based features. Green rows represent expression-based models with an analogous feature dimensionality to the correlation model. Blue rows represent expression-

based models with a feature dimensionality based on the number of biomarkers implicated in the correlation model. Table 4 reports these evaluation measures on the training dataset. The beta coefficients for each of the covariates (e.g. AJCC stage, correlation of *CD79* and *PCNA*) was learned on a 70% partition of the chemotherapy-naive patient cohort, and the prognostic ability of these covariates toward 5-year recurrence is evaluated here. Table 5 uses the same model parameters but on the remaining 30% partition of the cohort upon which these parameters weren't learned, and the predictive power against 5-year recurrence is again evaluated.

For the training data (Table 4) the correlation models have the best performance. Comparing correlation to expression (no clinical variables), model 2 has a higher concordance index on 5-year recurrence than both models 6 and 8 (7 expression features and 12 expression features). Model 2 also outperforms clinical features by a significant margin. Notably, the two expression models (6 and 8) do not outperform the clinical model 1. Adding clinical features to the correlation model (as in model 3) results in a performance boost. Clinical features also improve performance of both expression based models (7 vs. 6, and 9 vs. 8). The handpicked correlation models (4 and 5) slightly outperform the automated models (2 and 3) outlining that we can improve performance by examining correlation features in all subcellular compartments.

Similarly for the test data (Table 5), correlation features (model 2 and 4) outperform both clinical features (model 1) and expression features (model 6 and 8). Notably for the test data, expression features (model 6 and 8) perform worse than clinical features alone.

The feature sets learned in these 9 models are displayed below.

Correlation Features

1. Epithelial cyt-S6 vs. memb-CD163
2. Epithelial memb-NaKATPase vs. nuc-FOX03a
3. Stromal Claudin 1 vs. cMET epitomics
4. Epithelial memb-NaKATPase vs. cell-pGSK3beta
5. Epithelial cyt-panCK 26 vs. cyt-p4EBP1
6. Epithelial memb-CD79 vs. nuc-PCNA
7. Epithelial memb-Claudin 1 vs. cyt-Cleaved Caspase 3

Expression Features

1. Epithelial memb-CD8
2. Epithelial ratio cyt-EPCAM
3. Stromal nuc-CD3
4. Epithelial memb-COX2
5. Epithelial memb-CD3
6. Epithelial ratio memb-NaKATPase
7. Epithelial cyt-Beta Catenin
8. Epithelial ratio nuc-MSH2
9. Stromal nuc-CD8
10. Epithelial ratio nuc-P53
11. Epithelial ratio nuc-P4EBP1
12. Epithelial ratio cyt-Claudin 1

4.5.2 Patient stratification

Using the risk associated with the Youden's J statistic, we are able to partition the chemotherapy-naive patient cohort into low and high risk groups for the clinical, expression based, and correlation based feature sets (Fig 18). Comparing models 1, 2, 6, and 8 (clinical, correlation, and both expression sets), model 2 has the largest difference in recurrence between the low and high risk groups. This signifies that correlation is the best metric of the three for segregating patients into different groups as related to 5-year recurrence, potentially resulting in

Table 4: Performance metrics on training data for prognostic feature sets.

Model	C	R ²	HR	95% CI		P(HR)	P-val
1	0.754 ± 0.0184	0.0839 ± 0.0198	1.41	1.26	1.58	2e-07	6.2e-08
2	0.818 ± 0.0182	0.0883 ± 0.0199	1.15	1.11	1.2	7.21e-10	4.4e-13
3	0.832 ± 0.0144	0.13 ± 0.0316	1.1	1.08	1.13	4.44e-8	0
4	0.835 ± 0.0139	0.117 ± 0.0196	1.12	1.09	1.16	3.53e-12	6.51e-15
5	0.853 ± 0.0194	0.125 ± 0.0346	1.04	1.03	1.06	6.12e-7	0
6	0.755 ± 0.0205	0.0713 ± 0.013	1.23	1.15	1.31	1.15e-8	8.59e-11
7	0.805 ± 0.0216	0.11 ± 0.017	1.15	1.11	1.2	3.75e-13	0
8	0.75 ± 0.023	0.0587 ± 0.0142	1.11	1.07	1.14	3.62e-7	3.07e-12
9	0.805 ± 0.019	0.0873 ± 0.0257	1.09	1.06	1.11	2.25e-6	0

Concordance index, R-squared measure, hazards ratio, and confidence intervals are recorded for each of the nine prognostic models. The gray row signifies a model using only clinicopathology features, while red rows signify models containing correlation-based features. Green rows represent expression-based models with an analogous feature dimensionality to the correlation model. Blue rows represent expression-based models with a feature dimensionality based on the number of biomarkers implicated in the correlation model. These measure were computed on the testing partition of the patient cohort.

Table 5: Performance metrics on testing data for prognostic feature sets.

Model	C	R ²	HR	95% CI		P(HR)	P-val
1	0.746 ± 0.0472	0.081 ± 0.0366	1.53	1.23	1.91	0.00587	0.00485
2	0.792 ± 0.0451	0.0533 ± 0.0317	1.15	1.05	1.26	0.0302	0.0262
3	0.811 ± 0.0347	0.101 ± 0.0594	1.13	1.07	1.19	0.00159	0.000545
4	0.806 ± 0.041	0.0785 ± 0.056	1.11	1.06	1.17	0.00918	0.00616
5	0.805 ± 0.0536	0.109 ± 0.0496	1.05	1.03	1.07	0.00141	0.000585
6	0.702 ± 0.0571	0.0483 ± 0.0353	1.27	1.09	1.5	0.0826	0.0791
7	0.748 ± 0.0572	0.0442 ± 0.0247	1.13	1.05	1.22	0.0306	0.0249
8	0.673 ± 0.0607	0.0411 ± 0.024	1.13	1.03	1.24	0.024	0.0182
9	0.69 ± 0.0905	0.0283 ± 0.0233	1.08	1.01	1.16	0.174	0.166

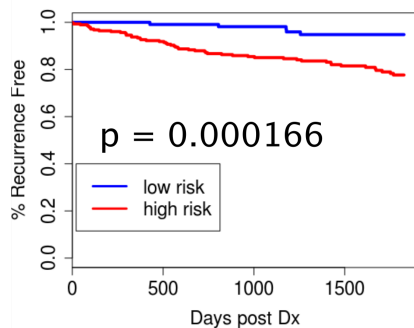
different strategies of care for the two groups. Additionally, correlation features improve the most upon the addition of clinical features (model 3), compared to the two expression-based models with clinical features (models 7 and 9). This demonstrates that correlation features are largely independent of the clinical features. All survival models reach significance. Additionally the ROC curve for model 2 is shown, where the black curve is the average ROC, and the other curves show the 20 individual runs.

4.5.3 Spatial organization of risk, via quadrant analysis and cell neighborhood analysis

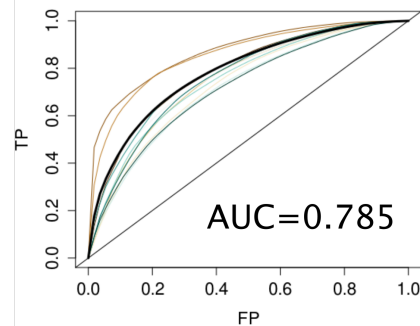
An important goal of this study is to examine how risk is spatially organized within tumor cores. The discover of high risk microdomains within tumor cores can potentially reveal tumor microenvironmental signatures associated with disease progression and CRC metastasis. Towards this end we have two strategies, computation of quadrant-based risk and computation of cell-level risk.

Fig. 19A represents the strategy associated with quadrant-based risk. The tumor core is split into four (or more) quadrants. Correlation features for each quadrant are then computed individually and the associated risk scores for each quadrant (from their respective Cox models) are binarized into low and high risk groups using the Youden’s J statistic. Then the correlation feature matrices can be further interrogated to identify biomarker correlations resulting in high risk scores. A GUI was developed to interrogate quadrant-based risk for a tumor core of interest. The tumor core is displayed in its four quadrants in this GUI. Each cell is colored by the distance between its local correlation measure (for the three most significant features, in the R,G,B channels respectively) and the mean correlation for that core, with epithelial cells denoted as circles and stromal cells denotes as squares. Core-level information such as name, AJCC stage, and 5-year recurrence status are displayed in a text box. Different commands allow for the epithelial, stromal, and total correlation matrix to be observed for each quadrant. Additionally, one can switch between showing the entire correlation matrix, and the 12 biomarkers associated with the 7 correlation feature model that makes risk-based decisions. Overall risk and quadrant level risk are also displayed in a

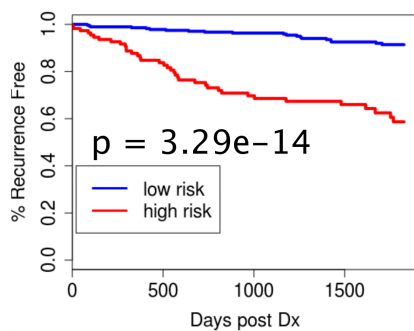
Model 1: Clincopathology



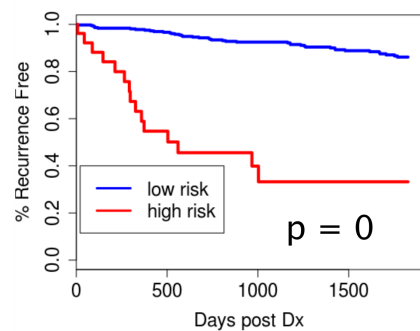
Model 2 ROC Curve



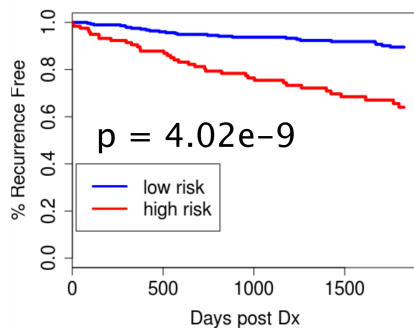
Model 2: Correlation (7 feature)



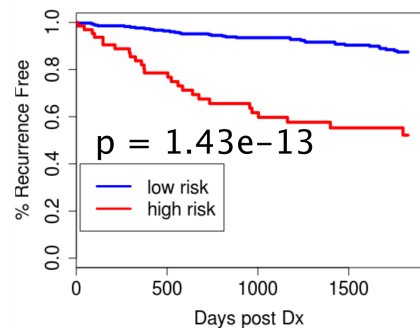
Model 3: Correlation+Clinical



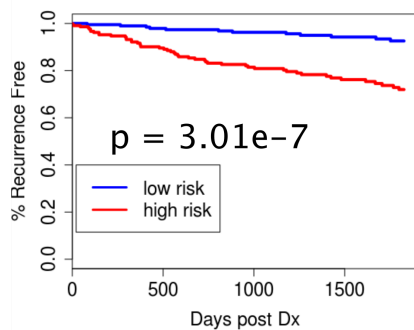
Model 6: Expression (7 feature)



Model 7: Expression (7)+Clinical



Model 8: Expression (12 feature)



Model 9: Expression (12)+Clinical

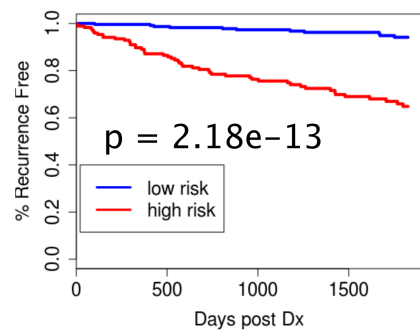


Figure 18: Kaplan-Meier Survival Curves for CRC 5-Year Recurrence Models.

text box (0 for low risk, 1 for high risk).

In Fig. 19A, AJCC stage III patient 216 in slide AGTA 269-3 undergoes 5-year recurrence, although only one quadrant (top left) is associated with high risk. The correlation matrices associated with the high risk quadrant and a low risk quadrant are displayed, with the rows and columns being represented by biomarkers of the top 7 correlation features as delineated in subsection 4.5.1. Notably, the high risk quadrant has a large stromal population that is relatively separate from the epithelial cell population. The low risk quadrant with the correlation matrix displayed has a notably lower stromal cell population. The remaining two low risk quadrant have large stromal populations but they are more intermixed with the epithelial cells.

Fig. 19B displays the results of cell-level risk scoring. Under this strategy, a correlation matrix is generated for each cell, based on a neighborhood of radius 200 pixels around the cell. All neighboring cells in that radius contribute to the correlation matrix, and thus the features that predict risk in the Cox model. A risk score is assigned to each cell in the tumor core. Risk values much greater than the low/high risk threshold are shown in bright red, risk values much less than the threshold are shown in bright green, and cells at the threshold are shown much dimmer, approaching black. In Fig. 19B, the leftmost tumor core is for slide AGA 260-4 patient 176, with AJCC stage II and 5-year recurrence. The middle core is slide AGA 260-3 patient 211, with AJCC stage I and no recurrence. The right core is slide AGTA 264-3 patient 131, with AJCC stage II and 5-year recurrence. Both examples exhibiting 5-year recurrence contain high risk cells contained to specific microdomains within the tumor core. The non recurring patient contains only a few high risk cells but contains several cells at the risk threshold that could potentially lead to higher risk if left unmonitored.

4.6 DISCUSSION

In this chapter we first described the current state of the art for colon cancer patient stratification through molecular subtyping, highlighting some of the gene signatures and pathway information that may be useful for monitoring disease progression. Then, using a MxIF

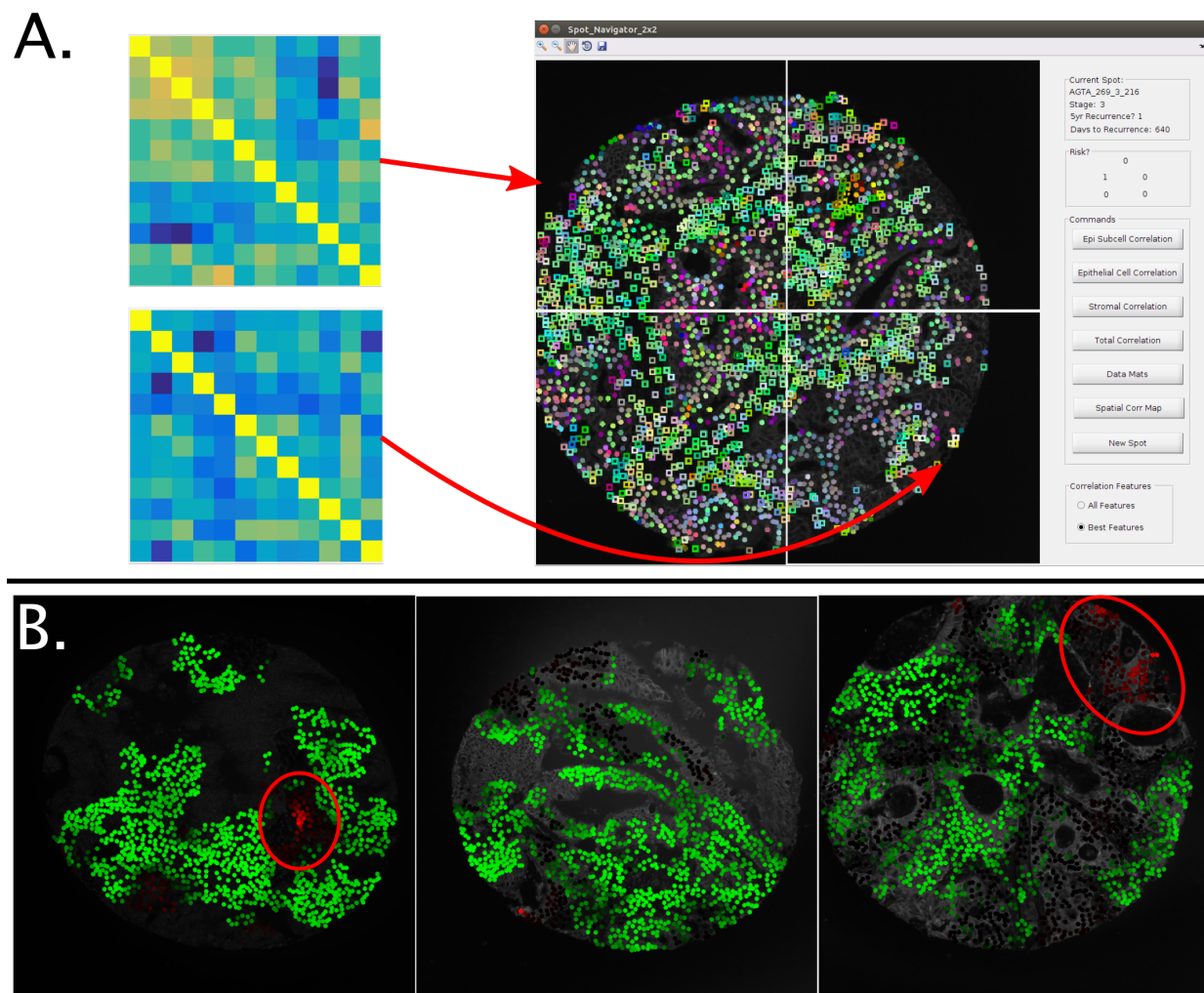


Figure 19: Spatial organization of risk using correlation features.

setup, we imaged a panel of protein biomarkers potential related to CRC biology and tumor microenvironment signaling. Biomarker expression feature and spatial correlation features were tested in univariate and multivariate models for predicting 5-year recurrence. A complete workflow for predicting clinical outcome, e.g. recurrence, survival, or metastasis, was described in the text and Fig. 17. These feature sets were also tested for efficacy in patient stratification, using Kaplan-Meier survival analysis.

Furthermore, we explored how risk was spatially organized within tumor cores, using quadrant-level analytics and cell-level analytics. We observed that risk typically localizes into microdomains, which we hypothesize may be regions of the tumor responsible for cancer progression, treatment resistance via cancer stem cell quiescence, and/or metastasis via development of a metastatic niche which can travel to distant sites much like circulating tumor cells.

There are several limitations of our analysis. Similar to genome-wide studies, we can only observe the set of biomarkers that we have chosen for the study a priori. However, unlike standard immunofluorescence studies, with multi- to hyper-plexed immunofluorescence we can simultaneously study many more covariates. Another issue is with the size of our tumor cores. To accurately study the effect of intratumor heterogeneity and the spatial distribution of risk, ideally we would be working with whole-slide tissue sections, or at least the largest microarray core size possible. We would like to capture as much of the potential tumor microenvironment as possible in our analyses. However, this study has adequately demonstrated the assembly of risk into spatial microdomains within the tumor core, which will only be ameliorated with larger data samples. The lack of randomization of AJCC staging across tissue microarray slides remains a large concern in our analysis, but many precautionary steps in the downstream analysis have been incorporated. However, we must retain some skepticism that slide-to-slide variation may account for some of the intratumor heterogeneity and variation across cancer staging that we observe.

There are many future directions that can be taken. For starters, there are many other feature selection routines that can be tested, to bring our correlation features down from 5940 correlation features and 385 expression features. Additionally, a dimensionality reduction step could always be added between the feature selection step and multivariate model step, as

is often done with genome wide association studies (although here the number of covariates is still much larger). A combination of expression and correlation features were tested in a combined model (not reported here), and although we did not achieve better results than with correlation alone, this could be explored further. For cell-based risk, more exhaustive testing could be done to find the ideal pixel radius for defining cell-cell contact. The increased performance of models 4 and 5 over models 2 and 3 signifies that even more prediction gains in terms of concordance index can be had, if we could exhaustively or intelligently try all combinations of biomarker correlations across all three subcellular components.

One goal that was originally in my thesis proposal was to test pointwise mutual information (PMI) and other diversity metrics, as described in Chapter 3, for prognostic potential (for recurrence, survival, etc.). However due to the expression level homogeneity, as shown in Fig. 14, we could not arrive at a solid cell phenotyping result upon which to compute cell phenotype co-occurrences. We did test a 2D PMI result using just epithelial and stromal labeling but the results did not reach significance. However, our collaboration with GE has also brought about supervised labeling software for phenotyping cells, which has been used and validated on a similar data set. This software would potentially allow for an accurate phenotyping step upon which PMI could be computed and tested against recurrence.

A very lofty future aim would involve samples of the primary neoplasm (which we have) as well as biopsies of metastases over the progression of the disease. This data would currently be expensive to obtain, and require at least five years of data collection, with more time needed to gain a sizable cohort of patient samples. However this would allow us to compare and contrast high risk microdomains in the primary neoplasm to elements of the metastatic niche in distant organs. This use of our methodology would allow for unprecedented studying of the tumor microenvironment of metastases, potentially pointing toward factors in the primary neoplasm that directly result in cancer recurrence and patient mortality. If this link was made, treatments could be developed to specifically counteract these high-risk microdomains, potentially to much success given the relative clonal stability of the tumor microenvironment constituents in relation to the malignant cells.

4.7 APPENDIX

4.7.1 Abbreviations

Presented in order of appearance:

CRC - colorectal cancer

FOLFIRI - adjuvant chemotherapy treatment of FOLFIRI: folinic acid, 5-fluorouracil, and irinotecan

FOLFOX -adjuvant chemotherapy treatment of FOLFIRI: folinic acid, 5-fluorouracil, and oxaliplatin

CTL - cytotoxic lymphocyte (CD8+ T cells)

TCGA - The Cancer Genome Atlas

CRCSC - colorectal cancer subtyping consortium

CMS -consensus molecular subtyping

CIMP - CpG island methylator phenotype

SCNA - somatic copy number alterations

CAF - cancer associated fibroblast

TAM - tumor associated macrophage

TIL - tumor infiltrating lymphocyte

EMT - epithelial-to-mesenchymal transistion

4.7.2 Colon cancer biomarkers

Table 6: Colon cancer biomarkers.

<i>pERK 1/2</i>	Mitogen activated protein kinase 1/2: central hub of mitogenic-progrowth signaling/this phosphorylation on the residues targeted here is indicative of kinase activity
<i>CD31</i>	Endothelial cell:cell junctions/immune cell transendothelial migration
<i>BetaCatenin</i>	Adherens junctions/Wnt signaling
<i>S6</i>	Ribosomal Protein S6: phosphorylation is associated with mitogens and growth factors and may regulate selective translation of particular classes of mRNAs defined by consensus sequences in untranslated regions
<i>pS6</i>	Ribosomal Protein S6: phosphorylation is associated with mitogens and growth factors and may regulate selective translation of particular classes of mRNAs defined by consensus sequences in untranslated regions.
<i>Fibronectin</i>	Secreted cell adhesion protein involved in basement membrane function
<i>BetaActin</i>	Microfilament protein: structure and motility
<i>CK1,5,6,8 (pck26)</i>	Pan-cytokeratin: recognizes most basal and luminal epithelial cell subtypes
<i>GLUT1</i>	Facilitated glucose/aldose uptake: upregulated in cancer cells/aerobic glycolysis/Warburg effect
<i>Na⁺/K⁺-ATPase</i>	Sodium potassium antiporter - ATP dependent: regulates membrane potential polarization/constitutive expression in all known cell types
<i>SMA</i>	Smooth muscle actin alpha: cytoskeletal protein of smooth muscle and vascular pericytes
<i>Albumin</i>	Extracellular transport: binds drugs and small metabolites
<i>CK19</i>	Cytokeratin 19: unpaired with basic cytokeratin
<i>EGFR</i>	Epidermal growth factor receptor: receptor tyrosine kinase/ binds epithelial growth factor/ homo or heterodimerization/ activates signal transduction
<i>p4EBP1</i>	Eukaryotic initiation factor 4E binding protein one: inhibits mRNA translation initiation/phosphorylation relieves inhibition of mRNA translation initiation
<i>WNT5a</i>	Wnt signaling secreted glycoprotein 5a: signals through canonical and non-canonical Wnt pathways/may affect cell motility and metastasis
<i>pNDRG1</i>	nMyc Downstream Regulated One: involved in cytoskeletal dynamics/adherens junctions/metabolism/poorly characterized/phosphorylation of this residue is downstream of mTORC2
<i>FOXO3a</i>	Forkhead Box-O3a: transcription factor involved in regulating intermediary metabolic enzymes and cell survival
<i>MLH1</i>	MutL homolog one: mismatch repair enzyme
<i>E-cadherin</i>	Epithelial specific homotypic adherens junctions/Wnt signaling/epithelial-mesenchymal transition/calcium dependent cell-cell adhesion/E7 integrin ligand
<i>pGSK3a</i>	Glycogen sythase kinase alpha: probable component of the destruction complex (-catenin degradation)/phsophorylation on the residue targeted here inactivates the kinase though the creation of a pseudosubstrate
<i>Lamin A/C</i>	Nuclear lamin protein: involved in organizing nuclear pores and chromatin
<i>pGSK3beta</i>	Glycogen sythase kinase beta: known component of the destruction complex (-catenin degradation)/phsophorylation on the residue targeted here inactivates the kinase though the creation of a pseudosubstrate

<i>EZH2</i>	Polycomb repressor complex two methyltransferase: inactivates gene expression through histone acetylation on H3K9 and H3K27
<i>ALDH1</i>	Cytosolic aldehyde dehydrogenase / retinoic acid metabolism / alcohol metabolism
<i>p21</i>	Cyclin dependent kinase inhibitor.
<i>Claudin1</i>	Epithelial and endothelial tight junction protein/epithelial barrier function/homo and heteropolymers and zona occludins protein binding. Neutrophil expression in CRC. Colocalizes with neutrophil specific elastase
<i>CD20</i>	B-lymphocyte protein of unknown function
<i>Indian Hedge Hog</i>	Hedgehog/pathcd/smoothened signaling: involved in regulating bone metabolism (ossification).
<i>pEGFR</i>	Epidermal growth factor receptor: receptor tyrosine kinase/ binds epithelial growth factor/ homo or heterodimerization/ activates signal transduction
<i>NDRG1</i>	nMyc Downstream Regulated One: involved in cytoskeletal dynamics/adherens junctions/metabolism/poorly characterized
<i>CD68</i>	Tissue macrophage scavenger receptor / endosomal-lysosomal glycoprotein / selectin-dependent migration
<i>TKLP1</i>	Transketolase-like protein 1: transfers 2 carbon ketol groups to aldose acceptor molecules (TKTL1)
<i>CD8a</i>	Antigen recognition/class I MHC binding/T-lymphocyte mediated killing
<i>CD79</i>	B-lymphocyte antigen receptor complex - mediates antigen dependent B-lymphocyte activation and signal transduction
<i>PTEN</i>	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN: tumor suppressor/opposes PI3K function by dephosphorylating the inositol 3'OH group/emerging nuclear functions related to DNA repair and apoptosis
<i>pMET</i>	Met Proto-Oncogene (Hepatocyte Growth Factor Receptor): instigates signaling through prosurvival and EMT signaling pathways/expression is required for gastrulation and deregulated overexpression is common in many cancers /Y1349 phosphorylation serves as a Gab1 binding site (a scaffold for activation of PI3K, PLC, and SHP2)
<i>pMAPKAPK2</i>	Mitogen-Activated Protein Kinase-Activated Protein Kinase 2: p38MAPK substrate motif/ active MAPKAPK2 stabilizes TNF and IL6 mRNA/destabilizes HSP27 complexes
<i>FOXO1</i>	Forkhead Box-O1: transcription factor involved in regulating intermediary metabolic enzymes and cell survival
<i>AKT</i>	Signal transduction:cell survival/anabolic metabolism
<i>CA9</i>	Membrane associated carbonic anhydrase: Hypoxia response/pH modulation
<i>Cleaved Caspase 3</i>	Cysteine peptidase - active form: activation of apoptosis/possible role in lipid metabolism through activation of SREBP
<i>ERK</i>	Mitogen activated protein kinase 1: central hub of mitogenic-progrowth signaling
<i>p-p38MAPK</i>	Mitogen-Activated Protein Kinase 14: integrates stress signals from environmental and cytokine stimuli/phosphorylation indicates the kinase is active.
<i>EPCAM</i>	Epithelial cell adhesion molecule: homotypic calcium independent cell adhesion molecule/ cell cycle modulation (myc, cyclin A and E)

<i>CD3e</i>	T-lymphocyte antigen recognition and signal transduction
<i>MSh2</i>	MutS homolog two: mismatch repair enzyme
<i>CD163</i>	Scavenger receptor cysteine rich family type B protein thought to be expressed in M2 polarized macrophages
<i>cMET- zenon, epitomics clone</i>	Met Proto-Oncogene (Hepatocyte Growth Factor Receptor): instigates signaling through prosurvival and EMT signaling pathways/expression is required for gastrulation and deregulated overexpression is common in many cancers
<i>PI3Kp110a</i>	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic Subunit Alpha: major lipid signaling transducer/activates PH domain proteins including Akt1/2/3/mutated in all large proportion of many solid tumor types
<i>4EBP1</i>	Eukaryotic initiation factor 4E binding protein one: inhibits mRNA translation initiation/phosphorylation relieves inhibition of mRNA translation initiation
<i>COX2</i>	Inducible prostaglandin-endoperoxide synthase: key enzyme in prostaglandin synthesis/inflammatory function
<i>PCNA</i>	Proliferating cell nuclear antigen: home trimeric clamp function feeds ssDNA to various polymerases and DNA repair enzymes
<i>p53</i>	Tumor protein 53: transcription factor involved in DNA damage responses resulting in repair and/or induction of apoptosis/ emerging functions in lipid metabolism through positive regulation of mevalonate pathway enzymes
<i>DAPI</i>	Nuclear counterstain - texture and intensity may indicate cell cycle/apoptosis/DNA dynamics
<i>Collagen IV</i>	Secreted basement membrane protein: type IV collagen network/angiogenesis inhibition/11 integrin ligand

4.8 ACKNOWLEDGEMENTS

This work was a collaboration with Shikhar Uttam, Peihong Zhu, Timothy R. Lezon, Albert Gough, Andrew M. Stern, Fiona Ginty, Brion Sarachan, D. Lansing Taylor and S. Chakra Chennubhotla. In addition to the Department of Computational and Systems Biology at Pitt, collaborators came from the Drug Discovery Institute. At GE Global Research in Niskayuna NY, collaborators came from the Biosciences Organization, as well as the Software Science and Analytics Organization.

5.0 CONCLUSIONS

As discussed in Chapter 1, spatial heterogeneity is a phenomenon that confounds consistent and effective cancer treatment, but is also overdue for closer scrutiny. Indeed, the composition of the tumor microenvironment and its spatial interactions beget heterocellular emergence of the malignant phenotype at the primary neoplasm, and may also be responsible for forming a hospitable niche at potential metastatic sites. While many different forms of analysis, especially genome wide association studies, have shown promise in characterizing intratumor heterogeneity, multiplexed immunofluorescence is uniquely positioned to model the spatial interactions between heterogeneous cell types through maintaining the tumor architecture. In this thesis, I have demonstrated that a characterization of spatial heterogeneity in the tumor microenvironment can have huge implications for better understanding the pathways that bring about cancers, and for predicting metastasis using only biopsy of the primary neoplasm.

In Chapter 2 we released a multiplexed image analysis tool, THRIVE, to assist in community-wide adoption of MxIF for heterogeneity analysis and other cancer-related studies. THRIVE contains an entire workflow for analyzing MxIF images and other high-dimensional images, from image segmentation and quantitation, to biomarker statistics and diversity scoring. THRIVE is open source, and its modular design allows researchers and developers to include their own algorithms within this workflow. Among other merits else, THRIVE will improve reproducibility of results across labs, allow for easy comparison of metrics and algorithms, and provide ease in the visualization of these types of images.

In Chapter 3, we described a method for quantifying spatial intratumor heterogeneity, using a standard breast cancer biomarker panel. Using pointwise mutual information as a diversity metric, we uncovered salient cell-cell cooccurrences, while accounting for the base-

line rarity of different cell phenotypes (analogous to the use of background distributions in genomic analyses). Heterogeneity maps can be used to discover cell-cell interaction motifs associated with different cancer subtypes.

In Chapter 4 we tested several biomarker expression features and spatial feature sets for predicting recurrence in colon cancer. We used spatial features to segregate patient cohorts into high risk and low risk regimes, potentially allowing clinicians to provide better care through patient stratification (especially in stage II CRC). Patient stratification is only the first step in the larger objective of personalized medicine. Additionally, we explored how patient risk is organized spatially with tumor core samples, paving the way for the discovery of microdomains within the tissue that are associated with metastatic potential.

Throughout my graduate career, I have been privileged to work with some of the best cancer researchers in the field, and gained early access to a promising and emerging imaging technology through my collaborations with GE Global Research. I believe that the work described in this thesis may provide a modest yet critical step toward a better understanding of the mechanisms that allow cancer to progress and eventually metastasize, typically leading to patient mortality [50, 63]. My work, and the work of future members of the Chennubhotla and Taylor labs, is just beginning.

BIBLIOGRAPHY

- [1] N. McGranahan and C. Swanton. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168(4):613–628, Feb 2017.
- [2] V. Almendro, A. Marusyk, and K. Polyak. Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol*, 8:277–302, 2013.
- [3] A. Marusyk, V. Almendro, and K. Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*, 12(5):323–34, 2012.
- [4] J. A. Joyce and J. W. Pollard. Microenvironmental regulation of metastasis. *Nat. Rev. Cancer*, 9(4):239–252, Apr 2009.
- [5] Douglas Hanahan and Robert A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.
- [6] T. A. Yap, M. Gerlinger, P. A. Futreal, L. Pusztai, and C. Swanton. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med*, 4(127):127ps10, 2012.
- [7] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–74, 2011.
- [8] J. E. Eckel-Passow, C. M. Lohse, Y. Sheinin, P. L. Crispen, C. J. Krco, and E. D. Kwon. Tissue microarrays: one size does not fit all. *Diagn Pathol*, 5:48, Jul 2010.
- [9] M. Zhang, A. V. Lee, and J. M. Rosen. The Cellular Origin and Evolution of Breast Cancer. *Cold Spring Harb Perspect Med*, 7(3), Mar 2017.
- [10] J. L. Fine. 21(st) century workflow: A proposal. *J Pathol Inform*, 5(1):44, 2014.
- [11] M. Janiszewska, L. Liu, V. Almendro, Y. Kuang, C. Paweletz, R. A. Sakr, B. Weigelt, A. B. Hanker, S. Chandarlapaty, T. A. King, J. S. Reis-Filho, C. L. Arteaga, S. Y. Park, F. Michor, and K. Polyak. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat. Genet.*, 47(10):1212–1219, Oct 2015.

- [12] M. Gerlinger, A. J. Rowan, S. Horswell, M. Math, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal, and C. Swanton. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.*, 366(10):883–892, Mar 2012.
- [13] R. Hoefflin, B. Lahrmann, G. Warsow, D. Hubschmann, C. Spath, B. Walter, X. Chen, L. Hofer, S. Macher-Goeppinger, Y. Tolstov, N. Korzeniewski, A. Duensing, C. Grulich, D. Jager, S. Perner, G. Schonberg, J. Nyarangi-Dix, S. Isaac, G. Hatiboglu, D. Teber, B. Hadaschik, S. Pahernik, W. Roth, R. Eils, M. Schlesner, H. Sultmann, M. Hohenfellner, N. Grabe, and S. Duensing. Spatial niche formation but not malignant progression is a driving force for intratumoural heterogeneity. *Nat Commun*, 7:ncomms11845, Jun 2016.
- [14] I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, and L. A. Garraway. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352(6282):189–196, Apr 2016.
- [15] S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, and B. E. Bernstein. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, 171(7):1611–1624, Dec 2017.
- [16] A. Kumar, E. A. Boyle, M. Tokita, A. M. Mikheev, M. C. Sanger, E. Girard, J. R. Silber, L. F. Gonzalez-Cuyar, J. B. Hiatt, A. Adey, C. Lee, J. O. Kitzman, D. E. Born, D. L. Silbergeld, J. M. Olson, R. C. Rostomily, and J. Shendure. Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. *Genome Biol.*, 15(12):530, Dec 2014.
- [17] R. Govindan. Cancer. Attack of the clones. *Science*, 346(6206):169–170, Oct 2014.
- [18] A. Bashashati, G. Ha, A. Tone, J. Ding, L. M. Prentice, A. Roth, J. Rosner, K. Shumansky, S. Kalloger, J. Senz, W. Yang, M. McConechy, N. Melnyk, M. Anglesio, M. T. Luk, K. Tse, T. Zeng, R. Moore, Y. Zhao, M. A. Marra, B. Gilks, S. Yip, D. G. Huntsman, J. N. McAlpine, and S. P. Shah. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.*, 231(1):21–34, Sep 2013.

- [19] A. G. Rivenbark, S. M. O'Connor, and W. B. Coleman. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *Am. J. Pathol.*, 183(4):1113–1124, Oct 2013.
- [20] Y. Sugihara, H. Taniguchi, R. Kushima, H. Tsuda, D. Kubota, H. Ichikawa, S. Fujita, and T. Kondo. Laser microdissection and two-dimensional difference gel electrophoresis reveal proteomic intra-tumor heterogeneity in colorectal cancer. *J Proteomics*, 78:134–147, Jan 2013.
- [21] B. Balluff, C. K. Frese, S. K. Maier, C. Schone, B. Kuster, M. Schmitt, M. Aubele, H. Hofler, A. M. Deelder, A. Heck, P. C. Hogendoorn, J. Morreau, A. F. Maarten Altelaar, A. Walch, and L. A. McDonnell. De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. *J. Pathol.*, 235(1):3–13, Jan 2015.
- [22] N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, and M. Wigler. Tumour evolution inferred by single-cell sequencing. *Nature*, 472(7341):90–94, Apr 2011.
- [23] Y. Wang, J. Waters, M. L. Leung, A. Unruh, W. Roh, X. Shi, K. Chen, P. Scheet, S. Vattathil, H. Liang, A. Multani, H. Zhang, R. Zhao, F. Michor, F. Meric-Bernstam, and N. E. Navin. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, 512(7513):155–160, Aug 2014.
- [24] A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev, and B. E. Bernstein. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190):1396–1401, Jun 2014.
- [25] A. A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, T. Danon, N. Perzov, and U. Alon. Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322(5907):1511–1516, Dec 2008.
- [26] J. M. Irish, R. Hovland, P. O. Krutzik, O. D. Perez, ?. Bruserud, B. T. Gjertsen, and G. P. Nolan. Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell*, 118(2):217–228, Jul 2004.
- [27] R. J. Critchley-Thorne, S. M. Miller, D. L. Taylor, and W. L. Lingle. Applications of cellular systems biology in breast cancer patient stratification and diagnostics. *Comb. Chem. High Throughput Screen.*, 12(9):860–869, Nov 2009.
- [28] R. L. Camp, G. G. Chung, and D. L. Rimm. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat. Med.*, 8(11):1323–1327, Nov 2002.

- [29] G. G. Chung, M. P. Zerkowski, S. Ghosh, R. L. Camp, and D. L. Rimm. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. *Lab. Invest.*, 87(7):662–669, Jul 2007.
- [30] M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R. J. Filkins, D. Hollman, V. Kamath, S. Kaanumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, C. C. McCulloch, E. McDonough, M. C. Montalto, Z. Pang, J. Rittscher, A. Santamaria-Pang, B. D. Sarachan, M. L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang, and F. Ginty. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci. U.S.A.*, 110(29):11982–11987, Jul 2013.
- [31] J. R. Lin. Multiplexed Single-Cell Imaging: Past, Present, and Future. *Assay Drug Dev Technol*, 15(1):8–10, Jan 2017.
- [32] G. M. Clarke, J. T. Zubovits, K. A. Shaikh, D. Wang, S. R. Dinn, A. D. Corwin, A. Santamaria-Pang, Q. Li, S. Nofech-Mozes, K. Liu, Z. Pang, R. J. Filkins, and M. J. Yaffe. A novel, automated technology for multiplex biomarker imaging and application to breast cancer. *Histopathology*, 64(2):242–255, Jan 2014.
- [33] D. A. Nelson, C. Manhardt, V. Kamath, Y. Sui, A. Santamaria-Pang, A. Can, M. Bello, A. Corwin, S. R. Dinn, M. Lazare, E. M. Gervais, S. J. Sequeira, S. B. Peters, F. Ginty, M. J. Gerdes, and M. Larsen. Quantitative single cell analysis of cell population dynamics during submandibular salivary gland development and differentiation. *Biol Open*, 2(5):439–447, May 2013.
- [34] M. T. Uhlik, J. Liu, B. L. Falcon, S. Iyer, J. Stewart, H. Celikkaya, M. O’Mahony, C. Sevinsky, C. Lowes, L. Douglass, C. Jeffries, D. Bodenmiller, S. Chintharlapalli, A. Fischl, D. Gerald, Q. Xue, J. Y. Lee, A. Santamaria-Pang, Y. Al-Kofahi, Y. Sui, K. Desai, T. Doman, A. Aggarwal, J. H. Carter, B. Pytowski, S. C. Jaminet, F. Ginty, A. Nasir, J. A. Nagy, H. F. Dvorak, and L. E. Benjamin. Stromal-Based Signatures for the Classification of Gastric Cancer. *Cancer Res.*, 76(9):2573–2586, 05 2016.
- [35] M. J. Gerdes, Y. Gokmen-Polar, Y. Sui, A. S. Pang, N. LaPlante, A. L. Harris, P. H. Tan, F. Ginty, and S. S. Badve. Single-cell heterogeneity in ductal carcinoma in situ of breast. *Mod. Pathol.*, Nov 2017.
- [36] Marcus Bode and Andreas Krusche. Toponome imaging system (tis): imaging the proteome with functional resolution. *Nature Methods— Application Notes*, 2007.
- [37] W. Schubert, B. Bonnekoh, A. J. Pommer, L. Philipsen, R. Bockelmann, Y. Malykh, H. Gollnick, M. Friedenberger, M. Bode, and A. W. Dress. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat. Biotechnol.*, 24(10):1270–1278, Oct 2006.

- [38] Walter Schubert, Andreas Dress, Mika Ruonala, Andreas Krusche, Reyk Hillert, Anne Gieseler, and Peter Walden. Imaging cyclus microscopy. *Proceedings of the National Academy of Sciences*, 111(2):E215–E215, 2014.
- [39] M. J. Gerdes. Reply to Schubert et al.: Regarding critique of highly multiplexed technologies. *Proc. Natl. Acad. Sci. U.S.A.*, 111(2):E216, Jan 2014.
- [40] W. Schubert. Advances in toponomics drug discovery: Imaging cyclus microscopy correctly predicts a therapy method of amyotrophic lateral sclerosis. *Cytometry A*, 87(8):696–703, Aug 2015.
- [41] M. Angelo, S. C. Bendall, R. Finck, M. B. Hale, C. Hitzman, A. D. Borowsky, R. M. Levenson, J. B. Lowe, S. D. Liu, S. Zhao, Y. Natkunam, and G. P. Nolan. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.*, 20(4):436–442, Apr 2014.
- [42] S. Rost, J. Giltneane, J. M. Bordeaux, C. Hitzman, H. Koeppen, and S. D. Liu. Multiplexed ion beam imaging analysis for quantitation of protein expression in cancer tissue sections. *Lab. Invest.*, 97(8):992–1003, Aug 2017.
- [43] J. R. Lin, M. Fallahi-Sichani, and P. K. Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat Commun*, 6:8390, Sep 2015.
- [44] K. Isse, A. Lesniak, K. Grama, B. Roysam, M. I. Minervini, and A. J. Demetris. Digital transplantation pathology: combining whole slide imaging, multiplex staining and automated image analysis. *Am. J. Transplant.*, 12(1):27–37, Jan 2012.
- [45] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*, 2:147–171, 2009.
- [46] C. J. Tape. The Heterocellular Emergence of Colorectal Cancer. *Trends Cancer*, 3(2):79–88, Feb 2017.
- [47] R. C. Jankowitz and A. V. Lee. The evolving role of multi-gene tests in breast cancer management. *Oncology (Williston Park, N.Y.)*, 27(3):210, 212, 214, Mar 2013.
- [48] S. Huang. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–3862, Dec 2009.
- [49] M. R. Junttila and F. J. de Sauvage. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*, 501(7467):346–354, Sep 2013.
- [50] D. M. Spagnolo, R. Gyanchandani, Y. Al-Kofahi, A. M. Stern, T. R. Lezon, A. Gough, D. E. Meyer, F. Ginty, B. Sarachan, J. Fine, A. V. Lee, D. L. Taylor, and S. C. Chennubhotla. Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. *J Pathol Inform*, 7:47, 2016.

- [51] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264, Sep 2015.
- [52] D. A. Sipkins, X. Wei, J. W. Wu, J. M. Runnels, D. Cote, T. K. Means, A. D. Luster, D. T. Scadden, and C. P. Lin. In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature*, 435(7044):969–973, Jun 2005.
- [53] Albert Gough, Timothy Lezon, James R. Faeder, Chakra Chennubhotla, Robert F. Murphy, Rebecca Critchley-Thorne, and D. Lansing Taylor. High-content analysis with cellular and tissue systems biology: a bridge between cancer cell biology and tissue-based diagnostics. In John Mendelsohn, Peter M. Howley, Mark A. Israel, Joe W. Gray, and Craig Thompson, editors, *The molecular basis of cancer*, book section 25, pages 369–392. Saunders/Elsevier, Philadelphia, PA, 4th edition, 2015.
- [54] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.*, 7(10):R100, 2006.
- [55] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, 9(7):676–682, Jun 2012.
- [56] P. Kankaanpää, L. Paavolainen, S. Tiitta, M. Karjalainen, J. Paivarinne, J. Nieminen, V. Marjomäki, J. Heino, and D. J. White. BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nat. Methods*, 9(7):683–689, Jun 2012.
- [57] T. Leinster and C. A. Cobbold. Measuring diversity: the importance of species similarity. *Ecology*, 93(3):477–489, Mar 2012.
- [58] Dirk Padfield, Jens Rittscher, and Badrinath Roysam. Spatio-temporal cell segmentation and tracking for automated screening. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 376–379. IEEE, 2008.
- [59] A. H. Gough, N. Chen, T. Y. Shun, T. R. Lezon, R. C. Boltz, C. E. Reese, J. Wagner, L. A. Verneti, J. R. Grandis, A. V. Lee, A. M. Stern, M. E. Schurdak, and D. L. Taylor. Identifying and quantifying heterogeneity in high content analysis: application of heterogeneity indices to drug discovery. *PLoS ONE*, 9(7):e102678, 2014.
- [60] Michal Aharon, Michael Elad, and Alfred Bruckstein. k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

- [61] S. M. Mahmoud, E. C. Paish, D. G. Powe, R. D. Macmillan, M. J. Grainge, A. H. Lee, I. O. Ellis, and A. R. Green. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.*, 29(15):1949–1955, May 2011.
- [62] T. Jacob, J. W. Gray, M. Troxell, and T. Q. Vu. Multiplexed imaging reveals heterogeneity of PI3K/MAPK network signaling in breast lesions of known PIK3CA genotype. *Breast Cancer Res. Treat.*, 159(3):575–583, Oct 2016.
- [63] D. M. Spagnolo, Y. Al-Kofahi, P. Zhu, T. R. Lezon, A. Gough, A. M. Stern, A. V. Lee, F. Ginty, B. Sarachan, D. L. Taylor, and S. C. Chennubhotla. Platform for quantitative evaluation of spatial intratumoral heterogeneity in multiplexed fluorescence images. *Cancer Res*, 77(21):e71–e74, 2017.
- [64] A. A. Alizadeh, V. Aranda, A. Bardelli, C. Blanpain, C. Bock, C. Borowski, C. Caldas, A. Califano, M. Doherty, M. Elsner, M. Esteller, R. Fitzgerald, J. O. Korbel, P. Lichter, C. E. Mason, N. Navin, D. Pe’er, K. Polyak, C. W. Roberts, L. Siu, A. Snyder, H. Stower, C. Swanton, R. G. Verhaak, J. C. Zenklusen, J. Zuber, and J. Zucman-Rossi. Toward understanding and exploiting tumor heterogeneity. *Nat. Med.*, 21(8):846–853, Aug 2015.
- [65] F. R. Balkwill, M. Capasso, and T. Hagemann. The tumor microenvironment at a glance. *J. Cell. Sci.*, 125(Pt 23):5591–5596, Dec 2012.
- [66] D. P. Tabassum and K. Polyak. Tumorigenesis: it takes a village. *Nat. Rev. Cancer*, 15(8):473–483, Aug 2015.
- [67] A. Marusyk, D. P. Tabassum, P. M. Altrock, V. Almendro, F. Michor, and K. Polyak. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520):54–58, Oct 2014.
- [68] V. Almendro, Y. K. Cheng, A. Randles, S. Itzkovitz, A. Marusyk, E. Ametller, X. Gonzalez-Farre, M. Munoz, H. G. Russnes, A. Helland, I. H. Rye, A. L. Borresen-Dale, R. Maruyama, A. van Oudenaarden, M. Dowsett, R. L. Jones, J. Reis-Filho, P. Gascon, M. Gonen, F. Michor, and K. Polyak. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep*, 6(3):514–527, Feb 2014.
- [69] A. Shirinifard, S. Thiagarajan, P. Vogel, and A. Sablauer. Detection of Phenotypic Alterations Using High-Content Analysis of Whole-Slide Images. *J. Histochem. Cytochem.*, 64(5):301–310, 05 2016.
- [70] R. Gyanchandani, Y. Lin, H. M. Lin, K. Cooper, D. P. Normolle, A. Brufsky, M. Fastuca, W. Crosson, S. Oesterreich, N. E. Davidson, R. Bhargava, D. J. Dabbs, and A. V. Lee. Intratumor Heterogeneity Affects Gene Expression Profile Test Prognostic Risk Stratification in Early Breast Cancer. *Clin. Cancer Res.*, 22(21):5362–5369, Nov 2016.

- [71] J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, J. Aach, K. Zhang, and G. M. Church. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc*, 10(3):442–458, Mar 2015.
- [72] C. Giesen, H. A. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. J. Schuffler, D. Grolimund, J. M. Buhmann, S. Brandt, Z. Varga, P. J. Wild, D. Gunther, and B. Bodenmiller. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods*, 11(4):417–422, Apr 2014.
- [73] M. Nederlof, S. Watanabe, B. Burnip, D. L. Taylor, and R. Critchley-Thorne. High-throughput profiling of tissue and tissue model microarrays: Combined transmitted light and 3-color fluorescence digital pathology. *J Pathol Inform*, 2:50, 2011.
- [74] E. C. Stack, C. Wang, K. A. Roman, and C. C. Hoyt. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods*, 70(1):46–58, Nov 2014.
- [75] C Radhakrishna Rao. Diversity and dissimilarity coefficients: a unified approach. *Theoretical population biology*, 21(1):24–43, 1982.
- [76] S. J. Potts, J. S. Krueger, N. D. Landis, D. A. Eberhard, G. D. Young, S. C. Schmechel, and H. Lange. Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. *Lab. Invest.*, 92(9):1342–1357, Sep 2012.
- [77] V. Almendro, H. J. Kim, Y. K. Cheng, M. Gonen, S. Itzkovitz, P. Argani, A. van Oudenaarden, S. Sukumar, F. Michor, and K. Polyak. Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res.*, 74(5):1338–1348, Mar 2014.
- [78] S. Y. Park, M. Gonen, H. J. Kim, F. Michor, and K. Polyak. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *J. Clin. Invest.*, 120(2):636–644, Feb 2010.
- [79] C. J. Rose, K. Naidoo, V. Clay, K. Linton, J. A. Radford, and R. J. Byers. A statistical framework for analyzing hypothesized interactions between cells imaged using multispectral microscopy and multiple immunohistochemical markers. *J Pathol Inform*, 4(Suppl):S4, 2013.
- [80] R. J. Steininger, S. Rajaram, L. Girard, J. D. Minna, L. F. Wu, and S. J. Altschuler. On comparing heterogeneity across biomarkers. *Cytometry A*, 87(6):558–567, Jun 2015.
- [81] Musodiq Bello, Ali Can, and Xiaodong Tao. Accurate registration and failure detection in tissue micro array images. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 368–371. IEEE, 2008.

- [82] F. Woolfe, M. Gerdes, M. Bello, X. Tao, and A. Can. Autofluorescence removal by non-negative matrix factorization. *IEEE Trans Image Process*, 20(4):1085–1093, Apr 2011.
- [83] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.
- [84] Alberto Santamaria-Pang, Yuchi Huangy, and Jens Rittscher. Cell segmentation and classification via unsupervised shape ranking. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 406–409. IEEE, 2013.
- [85] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng*, 57(4):841–852, Apr 2010.
- [86] Y. Al-Kofahi, W. Lassoued, K. Grama, S. K. Nath, J. Zhu, R. Oueslati, M. Feldman, W. M. Lee, and B. Roysam. Cell-based quantification of molecular biomarkers in histopathology specimens. *Histopathology*, 59(1):40–54, Jul 2011.
- [87] K. Francis and B. O. Palsson. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci. U.S.A.*, 94(23):12258–12262, Nov 1997.
- [88] T. D. Schneider, G. D. Stormo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, 188(3):415–431, Apr 1986.
- [89] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [90] Mark O Hill. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432, 1973.
- [91] Lou Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.
- [92] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal. Global cancer statistics, 2012. *CA Cancer J Clin*, 65(2):87–108, Mar 2015.
- [93] R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2016. *CA Cancer J Clin*, 66(1):7–30, 2016.
- [94] D. V. Tauriello, A. Calon, E. Lonardo, and E. Batlle. Determinants of metastatic competency in colorectal cancer. *Mol Oncol*, 11(1):97–119, Jan 2017.
- [95] L. Marisa, A. de Reynies, A. Duval, J. Selves, M. P. Gaub, L. Vescovo, M. C. Etienne-Grimaldi, R. Schiappa, D. Guenot, M. Ayadi, S. Kirzin, M. Chazal, J. F. Flejou, D. Benchimol, A. Berger, A. Lagarde, E. Pencreach, F. Piard, D. Elias, Y. Parc, S. Olschwang, G. Milano, P. Laurent-Puig, and V. Boige. Gene expression classification

- of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.*, 10(5):e1001453, 2013.
- [96] R. Gray, J. Barnwell, C. McConkey, R. K. Hills, N. S. Williams, and D. J. Kerr. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet*, 370(9604):2020–2029, Dec 2007.
- [97] N. Song, K. L. Pogue-Geile, P. G. Gavin, G. Yothers, S. R. Kim, N. L. Johnson, C. Lipchik, C. J. Allegra, N. J. Petrelli, M. J. O’Connell, N. Wolmark, and S. Paik. Clinical Outcome From Oxaliplatin Treatment in Stage II/III Colon Cancer According to Intrinsic Subtypes: Secondary Analysis of NSABP C-07/NRG Oncology Randomized Clinical Trial. *JAMA Oncol*, 2(9):1162–1169, Sep 2016.
- [98] D. M. Muzny, M. N. Bainbridge, K. Chang, H. H. Dinh, J. A. Drummond, G. Fowler, C. L. Kovar, L. R. Lewis, M. B. Morgan, I. F. Newsham, J. G. Reid, J. Santibanez, E. Shinbrot, L. R. Trevino, Y. Q. Wu, M. Wang, P. Gunaratne, L. A. Donehower, C. J. Creighton, D. A. Wheeler, R. A. Gibbs, M. S. Lawrence, D. Voet, R. Jing, K. Cibulskis, A. Sivachenko, P. Stojanov, A. McKenna, E. S. Lander, S. Gabriel, G. Getz, L. Ding, R. S. Fulton, D. C. Koboldt, T. Wylie, J. Walker, D. J. Dooling, L. Fulton, K. D. Delehaunty, C. C. Fronick, R. Demeter, E. R. Mardis, R. K. Wilson, A. Chu, H. J. Chun, A. J. Mungall, E. Pleasance, A. Robertson, D. Stoll, M. Balasundaram, I. Birol, Y. S. Butterfield, E. Chuah, R. J. Coope, N. Dhalla, R. Guin, C. Hirst, M. Hirst, R. A. Holt, D. Lee, H. I. Li, M. Mayo, R. A. Moore, J. E. Schein, J. R. Slobodan, A. Tam, N. Thiessen, R. Varhol, T. Zeng, Y. Zhao, S. J. Jones, M. A. Marra, A. J. Bass, A. H. Ramos, G. Saksena, A. D. Cherniack, S. E. Schumacher, B. Tabak, S. L. Carter, N. H. Pho, H. Nguyen, R. C. Onofrio, A. Crenshaw, K. Ardlie, R. Beroukhim, W. Winckler, G. Getz, M. Meyerson, A. Protopopov, J. Zhang, A. Hadjipanayis, E. Lee, R. Xi, L. Yang, X. Ren, H. Zhang, N. Sathiamoorthy, S. Shukla, P. C. Chen, P. Haseley, Y. Xiao, S. Lee, J. Seidman, L. Chin, P. J. Park, R. Kucherlapati, J. T. Auman, K. A. Hoadley, Y. Du, M. D. Wilkerson, Y. Shi, C. Liquori, S. Meng, L. Li, Y. J. Turman, M. D. Topal, D. Tan, S. Waring, E. Buda, J. Walsh, C. D. Jones, P. A. Mieczkowski, D. Singh, J. Wu, A. Gulabani, P. Dolina, T. Bodenheimer, A. P. Hoyle, J. V. Simons, M. Soloway, L. E. Mose, S. R. Jefferys, S. Balu, B. D. O’Connor, J. F. Prins, D. Y. Chiang, D. Hayes, C. M. Perou, T. Hinoue, D. J. Weisenberger, D. T. Maglinte, F. Pan, B. P. Berman, D. J. Van Den Berg, H. Shen, T. Triche, S. B. Baylin, P. W. Laird, G. Getz, M. Noble, D. Voet, G. Saksena, N. Gehlenborg, D. DiCara, J. Zhang, H. Zhang, C. J. Wu, S. Y. Liu, S. Shukla, M. S. Lawrence, L. Zhou, A. Sivachenko, P. Lin, P. Stojanov, R. Jing, R. W. Park, M. D. Nazaire, J. Robinson, H. Thorvaldsdottir, J. Mesirov, P. J. Park, L. Chin, V. Thorsson, S. M. Reynolds, B. Bernard, R. Kreisberg, J. Lin, L. Iype, R. Bressler, T. Erkkila, M. Gundapuneni, Y. Liu, A. Norberg, T. Robinson, D. Yang, W. Zhang, I. Shmulevich, J. J. de Ronde, N. Schultz, E. Cerami, G. Ciriello, A. P. Goldberg, B. Gross, A. Jacobsen, J. Gao, B. Kaczkowski, R. Sinha, B. Aksoy, Y. Antipin, B. Reva, R. Shen, B. S. Taylor, T. A. Chan, M. Ladanyi, C. Sander, R. Akbani, N. Zhang, B. M. Broom, T. Casasent, A. Unruh, C. Wakefield, S. R. Hamilton, R. Cason, K. A. Baggerly, J. N. Weinstein,

- D. Haussler, C. C. Benz, J. M. Stuart, S. C. Benz, J. Sanborn, C. J. Vaske, J. Zhu, C. Szeto, G. K. Scott, C. Yau, S. Ng, T. Goldstein, K. Ellrott, E. Collisson, A. E. Cozen, D. Zerbino, C. Wilks, B. Craft, P. Spellman, R. Penny, T. Shelton, M. Hatfield, S. Morris, P. Yena, C. Shelton, M. Sherman, J. Paulauskis, J. M. Gastier-Foster, J. Bowen, N. C. Ramirez, A. Black, R. Pyatt, L. Wise, P. White, M. Bertagnolli, J. Brown, T. A. Chan, G. C. Chu, C. Czerwinski, F. Denstman, R. Dhir, A. Dorner, C. S. Fuchs, J. G. Guillem, M. Iacocca, H. Juhl, A. Kaufman, B. Kohl, X. Van Le, M. C. Mariano, E. N. Medina, M. Meyers, G. M. Nash, P. B. Paty, N. Petrelli, B. Rabeno, W. G. Richards, D. Solit, P. Swanson, L. Temple, J. E. Tepper, R. Thorp, E. Vakiani, M. R. Weiser, J. E. Willis, G. Witkin, Z. Zeng, M. J. Zinner, C. Zornig, M. A. Jensen, R. Sfeir, A. B. Kahn, A. L. Chu, P. Kothiyal, Z. Wang, E. E. Snyder, J. Pontius, T. D. Pihl, B. Ayala, M. Backus, J. Walton, J. Whitmore, J. Baboud, D. L. Berton, M. C. Nicholls, D. Srinivasan, R. Raman, S. Girshik, P. A. Kigonya, S. Alonso, R. N. Sanbhadti, S. P. Barletta, J. M. Greene, D. A. Pot, K. R. Shaw, L. A. Dillon, K. Buetow, T. Davidsen, J. A. Demchok, G. Eley, M. Ferguson, P. Fielding, C. Schaefer, M. Sheth, L. Yang, M. S. Guyer, B. A. Ozenberger, J. D. Palchik, J. Peterson, H. J. Sofia, and E. Thomson. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–337, Jul 2012.
- [99] B. Perez-Villamil, A. Romera-Lopez, S. Hernandez-Prieto, G. Lopez-Campos, A. Calles, J. A. Lopez-Asenjo, J. Sanz-Ortega, C. Fernandez-Perez, J. Sastre, R. Alfonso, T. Caldes, F. Martin-Sanchez, and E. Diaz-Rubio. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. *BMC Cancer*, 12:260, Jun 2012.
- [100] A. Schlicker, G. Beran, C. M. Chresta, G. McWalter, A. Pritchard, S. Weston, S. Runswick, S. Davenport, K. Heathcote, D. A. Castro, G. Orphanides, T. French, and L. F. Wessels. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med Genomics*, 5:66, Dec 2012.
- [101] F. De Sousa E Melo, X. Wang, M. Jansen, E. Fessler, A. Trinh, L. P. de Rooij, J. H. de Jong, O. J. de Boer, R. van Leersum, M. F. Bijlsma, H. Rodermond, M. van der Heijden, C. J. van Noesel, J. B. Tuynman, E. Dekker, F. Markowitz, J. P. Medema, and L. Vermeulen. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat. Med.*, 19(5):614–618, May 2013.
- [102] A. Sadanandam, C. A. Lyssiotis, K. Homicsko, E. A. Collisson, W. J. Gibb, S. Wullschleger, L. C. Ostos, W. A. Lannon, C. Grotzinger, M. Del Rio, B. Lhermitte, A. B. Olshen, B. Wiedenmann, L. C. Cantley, J. W. Gray, and D. Hanahan. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.*, 19(5):619–625, May 2013.
- [103] E. Budinska, V. Popovici, S. Tejpar, G. D’Ario, N. Lapique, K. O. Sikora, A. F. Di Narzo, P. Yan, J. G. Hodgson, S. Weinrich, F. Bosman, A. Roth, and M. Delorenzi.

- Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J. Pathol.*, 231(1):63–76, Sep 2013.
- [104] P. Roepman, A. Schlicker, J. Tabernero, I. Majewski, S. Tian, V. Moreno, M. H. Snel, C. M. Chresta, R. Rosenberg, U. Nitsche, T. Macarulla, G. Capella, R. Salazar, G. Orphanides, L. F. Wessels, R. Bernards, and I. M. Simon. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int. J. Cancer*, 134(3):552–562, Feb 2014.
 - [105] A. Sadanandam, X. Wang, F. de Sousa E Melo, J. W. Gray, L. Vermeulen, D. Hanahan, and J. P. Medema. Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications. *Cell Cycle*, 13(3):353–357, 2014.
 - [106] J. Guinney, R. Dienstmann, X. Wang, A. de Reynies, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B. M. Bot, J. S. Morris, I. M. Simon, S. Gerster, E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G. C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J. W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S. H. Friend, P. Laurent-Puig, J. P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, and S. Tejpar. The consensus molecular subtypes of colorectal cancer. *Nat. Med.*, 21(11):1350–1356, Nov 2015.
 - [107] R. Dienstmann, L. Vermeulen, J. Guinney, S. Kopetz, S. Tejpar, and J. Tabernero. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer*, 17(2):79–92, 02 2017.
 - [108] R. Scherz-Shouval, S. Santagata, M. L. Mendillo, L. M. Sholl, I. Ben-Aharon, A. H. Beck, D. Dias-Santagata, M. Koeva, S. M. Stemmer, L. Whitesell, and S. Lindquist. The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. *Cell*, 158(3):564–578, Jul 2014.
 - [109] A. Calon, E. Espinet, S. Palomo-Ponce, D. V. Tauriello, M. Iglesias, M. V. Cespedes, M. Sevillano, C. Nadal, P. Jung, X. H. Zhang, D. Byrom, A. Riera, D. Rossell, R. Mangués, J. Massagué, E. Sancho, and E. Batlle. Dependency of colorectal cancer on a TGF β -driven program in stromal cells for metastasis initiation. *Cancer Cell*, 22(5):571–584, Nov 2012.
 - [110] A. Calon, E. Lonardo, A. Berenguer-Llargo, E. Espinet, X. Hernando-Mombalona, M. Iglesias, M. Sevillano, S. Palomo-Ponce, D. V. Tauriello, D. Byrom, C. Cortina, C. Morral, C. Barcelo, S. Tosi, A. Riera, C. S. Attolini, D. Rossell, E. Sancho, and E. Batlle. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.*, 47(4):320–329, Apr 2015.
 - [111] C. Isella, A. Terrasi, S. E. Bellomo, C. Petti, G. Galatola, A. Muratore, A. Mellano, R. Senetta, A. Cassenti, C. Sonetto, G. Inghirami, L. Trusolino, Z. Fekete, M. De Rid-

- der, P. Cassoni, G. Storme, A. Bertotti, and E. Medico. Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.*, 47(4):312–319, Apr 2015.
- [112] P. D. Dunne, D. G. McArt, C. A. Bradley, P. G. O’Reilly, H. L. Barrett, R. Cummins, T. O’Grady, K. Arthur, M. B. Loughrey, W. L. Allen, S. S. McDade, D. J. Waugh, P. W. Hamilton, D. B. Longley, E. W. Kay, P. G. Johnston, M. Lawler, M. Salto-Tellez, and S. Van Schaeybroeck. Challenging the Cancer Molecular Stratification Dogma: Intratumoral Heterogeneity Undermines Consensus Molecular Subtypes and Potential Diagnostic Value in Colorectal Cancer. *Clin. Cancer Res.*, 22(16):4095–4104, Aug 2016.
- [113] D. V. F. Tauriello and E. Batlle. Targeting the Microenvironment in Advanced Colorectal Cancer. *Trends Cancer*, 2(9):495–504, Sep 2016.
- [114] J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pages, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoue, P. Bruneval, P. H. Cugnenc, Z. Trajanoski, W. H. Fridman, and F. Pages. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, 313(5795):1960–1964, Sep 2006.
- [115] J. Galon, W. H. Fridman, and F. Pages. The adaptive immunologic microenvironment in colorectal cancer: a novel perspective. *Cancer Res.*, 67(5):1883–1886, Mar 2007.
- [116] Y. Naito, K. Saito, K. Shiiba, A. Ohuchi, K. Saigenji, H. Nagura, and H. Ohtani. CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer. *Cancer Res.*, 58(16):3491–3494, Aug 1998.
- [117] V. Deschoolmeester, M. Baay, F. Lardon, P. Pauwels, and M. Peeters. Immune Cells in Colorectal Cancer: Prognostic Relevance and Role of MSI. *Cancer Microenviron*, 4(3):377–392, Dec 2011.
- [118] C. C. Compton and F. L. Greene. The staging of colorectal cancer: 2004 and beyond. *CA Cancer J Clin*, 54(6):295–308, 2004.
- [119] Giovanni Fasano and Alberto Franceschini. A multidimensional version of the kolmogorov–smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1):155–170, 1987.
- [120] J. Galon, F. Pages, F. M. Marincola, H. K. Angell, M. Thurin, A. Lugli, I. Zlobec, A. Berger, C. Bifulco, G. Botti, F. Tatangelo, C. M. Britten, S. Kreiter, L. Chouchane, P. Delrio, H. Arndt, M. Asslaber, M. Maio, G. V. Masucci, M. Mihm, F. Vidal-Vanaclocha, J. P. Allison, S. Gnjjatic, L. Hakansson, C. Huber, H. Singh-Jasuja, C. Ottensmeier, H. Zwierzina, L. Laghi, F. Grizzi, P. S. Ohashi, P. A. Shaw, B. A. Clarke, B. G. Wouters, Y. Kawakami, S. Hazama, K. Okuno, E. Wang, J. O’Donnell-Tormey, C. Lagorce, G. Pawelec, M. I. Nishimura, R. Hawkins, R. Lapointe, A. Lundqvist, S. N. Khleif, S. Ogino, P. Gibbs, P. Waring, N. Sato, T. Torigoe, K. Itoh, P. S. Patel, S. N. Shukla, R. Palmqvist, I. D. Nagtegaal, Y. Wang, C. D’Arrigo, S. Kopetz, F. A.

- Sinicrope, G. Trinchieri, T. F. Gajewski, P. A. Ascierto, and B. A. Fox. Cancer classification using the Immunoscore: a worldwide task force. *J Transl Med*, 10:205, Oct 2012.
- [121] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15(4):361–387, Feb 1996.
- [122] H. C. Chen, R. L. Kodell, K. F. Cheng, and J. J. Chen. Assessment of performance of survival prediction models for cancer prognosis. *BMC Med Res Methodol*, 12:102, Jul 2012.